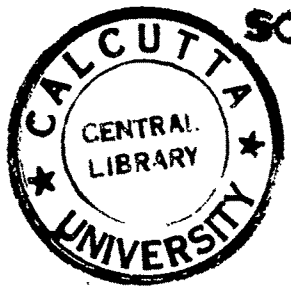
CU - A1774 - T1081

SOME PROBLEMS IN PROBABILITY THEORY

A thesis submitted to the University of Calcutta
for the degree of Doctor of Philosophy

By

R. RANGA RAO



SCANNED

Research and Training School
Indian Statistical Institute
Calcutta.

November, 1960.

SOME PROBLEMS IN PROBABILITY THEORY

PREFACE

This thesis is being submitted to the University of Calcutta in support of the author's application for the degree of Doctor of Philosophy (Arts). The thesis embodies research carried out by the author during the period 1958-60 under the supervision of Dr. R. R. Bahadur, Professor of Statistics at the Indian Statistical Institute, Calcutta.

The problems studied in this thesis are mainly concerned with limit theorems of probability theory. These problems have attracted the attention of many research workers in recent years. The principal content of the thesis is a solution of some of these problems.

The thesis consists of six chapters and two appendices. In chapter 1 the various problems considered in this thesis are described in general terms. The problems on the relation between weak convergence and uniform convergence of probability distributions on a general complete separable metric space are discussed in chapter 2, and their application to certain special spaces are given in chapter 3. In chapter 4 the convergence problems of sample distribution functions and their application to the laws of large numbers for general random functions are studied, and certain results extending and generalizing previous results of Mourier, and Fortet and Mourier are obtained. In chapters 5 and 6 the accuracy of the normal

approximation to the normed sum of independent and identically distributed random vectors is investigated. These chapters also contain a study of the asymptotic expansion of the 'error' in such approximations. In appendix A the properties of convex sets that are relevant to the thesis are discussed. In appendix B a generalization of the classical Euler-Maclaurin sum formula functions of several variables is given.

A few of the results of chapters 2-4 have appeared as an abstract in the Annals of Mathematical Statistics, 31 , 1960, p.241.

In concluding this preface, the author wishes to convey his sincere gratitude to Dr. R. R. Bahadur for his encouragement and interest in this work and for his valuable advice regarding the presentation of the results; and to Dr. C. R. Rao, Professor and Head of the department of Research and Training at the Indian Statistical Institute, for his constant encouragement.

The author also wishes to record his gratefulness to the Research and Training School, Indian Statistical Institute, for providing facilities for research. Thanks are also due to Mr. G. Das for his efficient typing of the thesis.

C O N T E N T S

		Page No.
Chapter 1.	Introduction	1
1.1.	General random functions	2
1.2.	Law of large numbers for general random functions	8
1.3.	On some questions of approximation related to the distribution of sums of independent random vectors.	13
1.4.	Summary of the Thesis.	21
Chapter 2.	On the relation between weak convergence and 'uniform' convergence of measures.	29
2.1.	Introduction	29
2.2.	Topological and measure theoretic preliminaries	29
2.3.	Weak convergence of measures	32
2.4.	Uniform convergence over a class of functions	36
2.5.	Uniform convergence over a class of sets	41
Chapter 3.	Application to finite dimensional Euclidean spaces and the space $C(0, 1]$.	48
3.1.	Introduction	48
3.2.	Some lemmas on measures on R_k	49
3.3.	Uniform convergence over the class $\mathcal{H}_m$	52
3.4.	Uniform convergence over the class $\mathcal{C}$ of all convex sets.	56
3.5.	Application to the convergence of continuous stochastic processes.	61

Chapter 4.	The law of large numbers for general random functions, and generalizations of the Glivenko-Cantelli lemma.		67
4.1.	Introduction		67
4.2.	Notation and definitions		68
4.3.	Convergence properties of sample distribution functions.		74
4.4.	Laws of large numbers for $\mathfrak{X}$-valued random variables.		80
4.5.	Laws of large numbers for $D[0, 1]$-valued random elements.		81
4.6.	Generalization of the Glivenko-Cantelli lemma		88
Chapter 5.	On the Central Limit Theorem in R_k. Estimation of the remainder term.		94
5.1.	Introduction		94
5.2.	General remarks about distributions in R_k.		97
5.3.	The main theorems.		106
5.4.	Expansions of the characteristic function of $Y^{(n)}$		111
5.5.	The functions $\tilde{P}_j(-\Phi)$.		116
5.6.	Some lemmas concerning $f_n(t)$.		119
5.7.	Some auxiliary functions		122
5.8.	Proof of Theorem 5.3.1.		129
5.9.	Proof of Theorem 5.3.2.		135

Chapter 6.	On the Central Limit Theorem in R_k. Asymptotic expansions of lattice distributions.	142
6.1.	Introduction	142
6.2.	Lattice distributions in R_k.	145
6.3.	Local limit theorems and estimation of the remainder.	149
6.4.	Asymptotic expansion of the probability $P_n(A)$ and estimation of the remainder term.	160
Appendix A.	On Convex Sets	167
A.1.	Introduction	167
A.2.	Inradius and minimum width of a convex set.	167
A.3.	Gauge function of a convex set and its properties.	170
A.4.	Blaschke's Theorem	177
A.5.	Convex sets and the Normal distribution in R_k.	180
A.6.	Convex sets and Fourier transforms	184.
Appendix B.	Generalization of the Euler-Maclaurin Sum formula to Functions of several variables.	189
B.1.	The functions $S_j(t)$.	189
B.2.	The Euler-Maclaurin sum formula for functions on R_1	191
B.3.	Generalization to functions of several variables	194
Bibliography	... ...	200

CHAPTER 1

Introduction

The problems of limiting behaviour which arise in probability theory may be divided roughly into three classes. (1) The study of limit theorems for distribution functions, where the underlying concept of convergence is 'weak convergence' ; (2) The limiting behaviour of random observations themselves or of functions of such observations; the problems of this type are known as strong limit problems; (3) Precise estimates in limit theorems, such as, estimation of the rapidity of convergence, or the asymptotic expansion of the 'error' etc. In accordance with the above classification this thesis falls naturally into three parts. The first part deals with certain general questions on the weak convergence of distributions in complete separable metric spaces. These questions are motivated in part by applications made in the second part, which part is devoted to the problem of proving strong law of large numbers for general random functions. In the third part we consider certain problems of approximation associated with the distribution function of sums of independent random vectors. We proceed now to discuss the contents of the thesis in greater detail.

2

1.1. General random functions: The study of random functions may be said to have begun with the work of Wiener on Brownian motion. In the decades following the appearance of Wiener's work this field has received a manysided development. For all this varied development the problem of limit theorems did not receive much attention till about 1950. An example of what could be achieved by considering limit theorems for random functions was given by Doob in 1949 ([12]). He showed that the classical theorem of Kolmogorov on the deviation between the true and empirical distribution function could be deduced from a limit theorem for random functions. The limit theorem which arises here is a sort of central limit theorem for random functions. A significant gap in Doob's heuristic argument was later filled by Donsker ([11]. 1952). It was thus made clear that limit theorems for random functions could lead to fruitful results.

One reason why consideration of limit theorems was delayed was that a random function was usually regarded as an aggregate of random variables. This point of view has tended to obscure the problems of limiting behaviour underlying a random function. However, the way was cleared with the introduction of abstract random element. i.e. a random mapping into an abstract space. In most of the examples met with in practice, the abstract space X where the random element is realised, is a function space and there is a natural topology at hand. It is thus natural to consider random elements into topological spaces, in particular metric spaces. In other words if X denotes

an arbitrary metric space and $\mathcal{B}$ the class of all Borel subsets of X, then any measurable map of a probability space (Ω, S, P) into X will be called an X-valued random variable (or a random element in X).

Basic to the study of limit theorems is the study of the notion of convergence for probability distributions. For distributions on the real line these studies are all classical. For general random functions or random variables taking values in a (separable) metric space it is necessary to generalize the concept of convergence suitably. If ξ_n is a sequence of r.f, a knowledge of the convergence of the distribution function of the random variable $F(\xi_n)$, a functional of the process ξ_n, is often of importance. Consequently a convergence definition for r.f. is adopted which ensures the convergence of distributions of the random variables $F(\xi_n)$ to that of the limiting process, to as large a class of functions as possible. This is achieved by the definition: $P_n \Rightarrow P$ (converges weakly) if and only if the distribution of $F(\xi_n)$ converges to that of $F(\xi)$, for every continuous functional F on X. An equivalent definition is, $P_n \Rightarrow P$, if and only if

$$\int f \, d P_n \to \int f \, d P$$

for each bounded continuous function f on X. This convergence has been investigated by Alexandrov in his memoirs in Mat. Sbornik ([1], [ ], [ ]; particularly chapter 4). A study of this

convergence with special reference to applications to probability theory was carried out by Prohorov (1956, [29]).

The first part of this thesis considers two problems which arise directly from this definition of weak convergence or its equivalent forms. The first one may be described as follows. Suppose μ_n, μ is a sequence of distributions such that $\mu_n \Rightarrow \mu$. The problem then is to find classes $\mathcal{A}$ of continuous functions such that

$$\sup_{f \in \mathcal{A}} \left| \int f \, d\mu_n - \int f \, d\mu \right| \to 0 \tag{1.1.1}$$

as $n \to \infty$. The motivation and interest of this problem lies in its applications. It provides an alternative approach to the solution of certain strong limit theorems for random functions. This will be described later when we deal with strong limit theorems. Though several particular cases of (1.1.1) are known, this problem in this generality does not seem to have received attention in the literature. For instance let X be the set of positive integers with the discrete topology. Then, as is well known, $\mu_n \Rightarrow \mu$, if and only if

$$\| \mu_n - \mu \| \to 0$$

where $\| \ \|$ denotes the norm of a signed measure. This is an assertion of type (1.1.1) with $\mathcal{A}$ = [all functions f with

$| f(x) | \leq 1$ for all $x \in X$]. Another such result occurs in a disguised form in Wald ([38], Chapter, Theorem 2.14).

The second problem mentioned above is an analogue of the first one. A well known characterization of weak convergence states that μ_n weakly converges to μ when and only when $\mu_n(A) \to \mu(A)$ for each continuity set A of μ. If $\mathcal{B}(\mu)$ denotes the class of continuity sets of μ, then the problem is to isolate subclasses $\mathcal{A} \subset \mathcal{B}(\mu)$ over which the convergence $\mu_n(A) \to \mu(A)$ can be asserted to be uniform i.e.

$$(1.1.2) \qquad \sup_{A \in \mathcal{A}} | \mu_n(A) - \mu(A) | \to 0$$

as $n \to \infty$. As an instance of a known result of this type we may mention the theorem of Polya which states that if $F_n(x)$ is a sequence of distributions on the real line converging weakly to a continuous distribution $F(x)$, then

$$\sup_{-\infty < x < \infty} | F_n(x) - F(x) | \to 0$$

as $n \to \infty$.

This problem is also of some interest in applications. We will present two examples where questions of this type arise naturally.

Let ξ_n be a sequence of k-dimensional random vectors with means zero and variance covariance matrix = C_n (nonsingular). The problem then arises of discussing the asymptotic behaviour of ξ_n

when it is known that

$$(1.1.3) \qquad P\left\{ \frac{\underset{\sim}{\lambda}' \cdot \underset{\sim}{\xi}_n}{(\underset{\sim}{\lambda}' C_n \underset{\sim}{\lambda})^{\frac{1}{2}}} \leq x \right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2} dx$$

as $n \to \infty$. The asymptotic behaviour of $\underset{\sim}{\xi}_n$ is easily found when $C_n = C$ for all n, or $C_n \to C$. In this case it is not difficult to show that $\underset{\sim}{\xi}_n$ converges weakly to a normal distribution with mean zero and variance covariance matrix C. The interest therefore shifts to the case when C_n does not have any limit. In this case one can still ask whether $\underset{\sim}{\xi}_n$ is asymptotically equivalent to a normal variable with mean zero and variance covariance matrix C_n in the sense that

$$(1.1.4) \qquad P[\, \underset{\sim}{\xi}_n \in R \,] - \int_R d\, N(0, C_n) \to 0$$

as $n \to \infty$, where R denotes an arbitrary rectangle in E_k^*. The answer is in the affirmative and the solution essentially reduces to an assertion of type (1.1.2). To see this, let B_n be a non singular matrix such that $B_n' C_n B_n = I$ (unit matrix) and let $\underset{\sim}{\eta}_n = B_n \underset{\sim}{\xi}_n$. Then (1.1.3) implies that

$$P\left\{ \frac{\underset{\sim}{\lambda}' \cdot \underset{\sim}{\eta}_n}{(\underset{\sim}{\lambda}' \cdot \underset{\sim}{\lambda})^{\frac{1}{2}}} \leq x \right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2} dx$$

---

* The author is indebted to S.K.Mitra of the Indian Statistical Institute for drawing his attention to the above problem.

Thus η_n converges weakly to N(0, I). If we now interpret (1.1.4) in terms of ξ_n it is equivalent to

$$(1.1.5) \qquad P[\eta_n \varepsilon B_n(R)] - \int_{B_n(R)} d\,N(0, I) \to 0$$

as $n \to \infty$, where the set $B_n(R)$ denotes the image of R under the transformation B_n. But (1.1.5) is an immediate consequence of the statement

$$\sup_A \left| P[\eta_n \varepsilon A] - \int_A d\,N(0, I) \right| \to 0$$

as $n \to \infty$, where the supremum is taken over all k-sided polyhedrons. (see Theorem 3.3.1 of chapter 3). This example makes it clear that a solution of the second problem could be of use in studying the asymptotic behaviour of a sequence of random elements, when actual convergence in distribution fails to take place.

This question seems to be also of some interest in problems of approximation which arise in probability theory. Suppose for instance that a certain sequence of distributions μ_n is known to converge weakly to a distribution μ. A refinement of this limit theorem often useful in applications would be to obtain a precise estimate of the difference $\mu_n(A) - \mu(A)$ for certain important class of sets A. In such problems the assertion (1.1.2) points out classes of sets over which one can expect to find uniform bounds for the error

$\mu_n(A) \to \mu(A)$. An example of this type is provided by the central limit theorem for random vectors in k-dimensions. This will be discussed in greater detail later on.

As regards weak convergence, the contribution of this thesis consists in giving a fairly general solution to the two problems mentioned above, together with a detailed analysis in the special case of Euclidean spaces.

1.2. Law of large numbers for general random elements. A problem of great theoretical importance with regard to a sequence of random elements $\xi_1, \xi_2, \ldots$ is the so called law of large numbers (LLN). A LLN is typically a theorem which guarentees the convergence of the average

$$S_n = \frac{\xi_1 + \cdots + \xi_n}{n}$$

to a limit in some sense. A classical result of this type is Birkhoff's individual ergodic theorem (Halmos [19], p. 18). This states that if $\xi_1, \xi_2, \ldots \xi_n, \ldots$ is a strictly stationary sequence of real valued random variables and $E \mid \xi_i \mid < \infty$, then there exists a random variable Y such that

$$P [S_n \to Y] = 1.$$

This is one of the important results in probability theory and it is natural that its extension to the more general situation of random

functions should be studied. That is the main object in the second part of the thesis.

In generalizing this result to random functions it is clear that it is appropriate to consider random variables taking values in a vector space. One of the simplest of these is that of a separable Banach space and this situation has been considered by Mourier ([28]). Mourier shows that if $\{\xi_n\}$ is a strictly stationary sequence of random elements in a Banach space X such that $E\|\xi_i\| < \infty$ (here $\|\cdot\|$ denotes the norm in X), then there exists a random element ξ in X such that

$$P[\|\eta_n - E\xi_1\| \to 0] = 1$$

where $\eta_n = n^{-1}(\xi_1 + \ldots + \xi_n)$. $E\xi_1$ is the expected value of ξ_1[1]. She also obtained a result on the convergence of η_n in the mean of order α. i.e. if $E\|\xi_i\|^\alpha < \infty$ $(\alpha \geq 1)$, then

$$\lim_{n \to \infty} E\|\eta_n - E\xi_1\|^\alpha = 0$$

Though these results have many applications, the assumption of separability, which is crucial in her methods, excludes many interesting cases. An example of this is provided by processes realised in D[0, 1] - the set of all functions on the interval [0, 1] with discontinuities if any, only of the first kind. More precisely, D[0, 1] is the family of all functions $x(t)$ on $0 \leq t \leq 1$ such that

1) $E\xi_1$ is an element of X determined by the condition that $(E\xi_1, x^*) = E(\xi_1, x^*)$, for each $x^* \in X^*$. For its existence and other details see Hille and Phillips ([20], Chapter III) and chapter 4 of this thesis.

10

$$(1)\quad \lim_{t \downarrow 0} x(t) = x(0); \quad \lim_{t \uparrow 1} x(t) = x(1),$$

$$(2)\quad \lim_{s \downarrow t} x(s) \text{ and } \lim_{s \uparrow t} x(s)$$

both exist, and

$$(3)\quad \lim_{s \downarrow t} x(s) = x(t)$$

It may be remarked that processes that are realised in $D[0, 1]$ are of sufficiently frequent occurrence to merit detailed study. However, the natural metric for $D[0, 1]$ - namely the sup norm

$$\| x \| = \sup_{0 \le t \le 1} | x(t) |$$

converts it into a non-separable Banach space.

In this connection, it may also be pointed out that the classical Glivenko - Cantelli lemma is a sort of law of large numbers for random elements into $D[0, 1]$*. For suppose that $\xi_1, \xi_2, \ldots$ is a sequence of independent repetitions of a random variable with the uniform distribution in $[0, 1]$, then let $X_1, X_2, \ldots$ be random elements in $D[0, 1]$ defined as follows:

---

* Here and in the sequel by a random element in $D[0, 1]$ is meant a function on a probability space (Ω, S, P) into $D[0, 1]$ say $x(t, w)$ such that for each $t \in [0, 1]$, $x(t, w)$ is a random variable.

11 $$X_n(t) = u_{\xi_n}(t) \qquad \text{for } 0 \le t \le 1$$

$$u_\tau(t) = \begin{cases} 0 & \text{if } \tau < t \\ 1 & \text{if } \tau \ge t \end{cases}$$

Then it is clear that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random elements in $D[0, 1]$. Now the Glivenko Cantelli theorem states that

$$P\left[\lim_{n \to \infty} \sup_{0 \le t \le 1} | F_n(t) - t | = 0\right] = 1$$

where $F_n(t)$ is the empirical distribution function based on the sample $\xi_1, \ldots \xi_n$. Since

$$F_n(t) = n^{-1}(X_1 + \ldots + X_n)$$

$$E\, X_1(t) = t$$

this could also be written as

$$P\left[\lim_{n \to \infty} \sup_{0 \le t \le 1} \left| \frac{X_1(t) + \ldots + X_n(t)}{n} - E\, X_1(t) \right| = 0\right] = 1.$$

It is clear that this is a form of strong law of large numbers for the sequence of random functions $\{X_n(t)\}$. This was pointed out by Gnedenko (1958, [17]) who raised the problem of extending the strong law of large numbers to $D[0, 1]$ valued random variables. It may be pointed out that even in this case the random function $X_1(t)$ defined

12 above is not concentrated in any separable subset of $D[0, 1]$ (under the sup norm). Thus a genuine extension of Mourier's results will be necessary to cover even this simplest case.

In this thesis we present a solution of this problem. (cf. Theorem 4.5.2 of chapter 4), employing methods different from those of Mourier. It is based on the following observation. If μ_n denotes the empirical distribution based on the sample $\xi_1, \ldots, \xi_n$ of random functions (i.e. a measure with masses $\frac{1}{n}$ at the points $\xi_1, \ldots, \xi_n$) and μ their common distribution then the strong law of large numbers often takes the following the form:

$$\sup_{f \in \mathcal{Q}} \left| \int f \, d\mu_n - \int f \, d\mu \right| =$$

$$= \sup_{f \in \mathcal{Q}} \left| \frac{f(\xi_1) + \ldots + f(\xi_n)}{n} - E f(\xi_1) \right|$$

tends to zero with probability one, where $\mathcal{Q}$ is a suitable class of functionals on X. Thus for instance when ξ_i's take values in a separable Banach space, $\mathcal{Q}$ consists of all linear functionals with norm ≤ 1. The problem therefore reduces to an investigation of uniform convergence of the sequence of empirical distributions over a class of functions. A study of convergence properties of empirical d.f.s is made in this part of the thesis.

1.3. On some questions of approximation related to the distribution of sums of independent random vectors. The increasing complexity of distributions that one comes across in various statistical procedures has once again emphasized the need to find usable approximations to distributions. This is particularly so with regard to the large sample theory of statistical inference. Here the problems take the form of obtaining refinements to limit theorems by actual estimations of the errors of approximation. A classical case is afforded by the central limit theorem. Indeed a large body of the literature on approximation in probability theory centres around this case.

Let us first consider the case of real valued random variables. Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be a sequence of independent and identically distributed random variables with $E\xi_1 = 0$ and $E\xi_1^2 = \sigma^2$. Let $F_n(x)$ denote the distribution function of the normed sum

$$\frac{\xi_1 + \ldots + \xi_n}{\sigma\sqrt{n}}$$

Then as is well known $\lim_{n \to \infty} F_n(x) = \Phi(x)$ for each x where Φ is the standard normal distribution with mean zero and variance unity. In connection with this theorem the following two problems arise naturally.

1. To find an estimate of the error $= |F_n(x) - \Phi(x)|$.

2. To find an asymptotic expansion for the difference $F_n(x) - \Phi(x)$.

14 Both these problems has been the subject of investigation by many writers notably Chebyshev, Lyapounov, Charlier, Edgeworth and later by Cramér, Berry, Esseen, and others.

A solution to the first problem was given in a definitive form by Berry [5] and Esseen [15], as follows. Suppose $E|\xi_1|^3 < \infty$, then

$$(1.3.1) \qquad |F_n(x) - \Phi(x)| < C \frac{E|\xi_1|^3}{\sigma^3} \frac{1}{n^{\frac{1}{2}}}$$

for all x, where C is a universal constant. (According to Wallace [39], C can be taken to be 2.05). Since the error in (1.3.1) is of the order of $n^{-\frac{1}{2}}$ the approximation is good only for very large samples. This situation at once leads us to the second problem of obtaining an asymptotic expansion of the difference $F_n - \Phi$.

Formal expansions of $F_n - \Phi$ in terms of Φ and its derivatives have been given by Chebyshev, Charlier, Bruns, Edgeworth and others. It is however preferrable to consider the rearranged series which can be written in the form

$$(1.3.2) \qquad F_n(x) = \Phi(x) + e^{-\frac{1}{2}x^2}\left\{ n^{-\frac{1}{2}} p_1(x) + n^{-1} p_2(x) + \dots \right\}$$

where $p_1(x)$, $p_2(x)$ etc., are certain polynomials in x. This expansion is commonly known as the Edgeworth expansion. Another

15 fruitful way of viewing the expansion (1.3.2) is to consider the formal expansion of the characteristic function $f_n(t)$ of $F_n(x)$. Let $f(t)$ denote the characteristic function of F and $\lambda_2, \lambda_3, \ldots$ be the cumulants of F. Then clearly

$$[\, f(t/\sqrt{n})]^n = e^{-\frac{1}{2}t^2} \exp\Big[\sum_{r=3}^{\infty} \frac{\lambda_r}{r!} (it)^r n^{-\frac{1}{2}(r-2)} \Big]$$

where we have supposed that $\lambda_2 = \sigma^2 = 1$. Let

$$\exp\Big[\sum_{r=3}^{\infty} n^{-\frac{1}{2}(r-2)} \frac{\lambda_r}{r!} (it)^r \Big] = 1 + \sum_{j=1}^{\infty} n^{-\frac{1}{2}j} P_j(it)$$

where $P_j(w)$ is a polynomial in w. Thus the formal expansion of ch.f of F_n is the following:

$$(1.3.2) \qquad f_n(t) = e^{-\frac{1}{2}t^2} + \sum_{j=1}^{\infty} n^{-\frac{1}{2}j} P_j(it)\, e^{-\frac{1}{2}t^2}$$

The relationship between the two expansions (1.3.1) and (1.3.2) is a simple one. It is that the function $p_j(x)\, e^{-\frac{1}{2}x^2}$ has $P_j(it)\, e^{-\frac{1}{2}t^2}$ as its Fourier transform, so that the right hand side of (1.3.2) is the Fourier transform of the expansion in (1.3.1).

Cramér ([10], chapter 7) investigated the order of the remainders in the Edgeworth expansion for distributions satisfying the following condition:

16

(C) $$\lim_{|t| \to \infty} |f(t)| < 1$$

which is for instance satisfied if F has an absolutely continuous component. More precisely, Cramér showed that when (C) is fulfilled,

$$(1.3.3) \quad F_n(x) = 1 + \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} p_j(x) e^{-\frac{1}{2}x^2} + O(n^{-\frac{1}{2}(s-2)}).$$

The condition (C) is however not satisfied for discrete distributions and the expansion (1.3.3) is then not generally valid.

Esseen studied the expansion problem when the distribution F of ξ_1 is of lattice type, i.e. is confined to a set of equally spaced points. In this case he showed that an expansion of the following form is always valid:

$$(1.3.4) \quad F_n(x) = \Phi(x) + \frac{p_1(x)e^{-\frac{1}{2}x^2}}{\sqrt{n}} + \frac{S_1(x)e^{-\frac{1}{2}x^2}}{\sqrt{n}} + O\left(\frac{1}{n}\right)$$

where $S_1(x)$ is a certain discontinuous function depending on the lattice. These are some of the important results known for real valued random variables. An excellent exposition of these results may be found in Esseen ([15]), or in the book of Gnendenko and Kolmogorov [18].

Let us now consider these problems for random vectors in R_k. Let $Z^{(1)}, Z^{(2)}, \ldots$ be a sequence of independent and identically

17 distributed random vectors with mean vector zero and variance covariance matrix I. If $F_n(\underset{\sim}{x})$ denotes the distribution function of

$$\frac{\underset{\sim}{Z}^{(1)} + \ldots + \underset{\sim}{Z}^{(n)}}{\sqrt{n}},$$

and Φ the standard normal distribution function, then, as is well-known,

$$\lim_{n \to \infty} F_n(\underset{\sim}{x}) = \Phi(\underset{\sim}{x})$$

for each vector $\underset{\sim}{x}$. Evidently, the two problems mentioned earlier in the case $k = 1$ are also meaningful here, and call for solution. However, for random vectors in R_k it is preferable to consider the probability functions $P_n(A)$ and $\Phi(A)$ where

$$P_n(A) = P\left\{\frac{\underset{\sim}{Z}^{(1)} + \ldots + \underset{\sim}{Z}^{(n)}}{\sqrt{n}} \in A\right\}$$

and

$$\Phi(A) = \int_A (2\pi)^{-\frac{1}{2}k} \exp\left[-\tfrac{1}{2}(x_1^2 + \ldots + x_k^2)\right] d\underset{\sim}{x}.$$

This is because the distribution function $F_n(x_1, \ldots, x_k)$ is associated with a coordinatization in R_k and the coordinate system in R_k itself can be arbitrary. Also there are sets A other than rectangles for which it would be interesting and important to study the error $P_n(A) - \Phi(A)$. Consequently, the two problems (1) estimations of the error $|P_n(A) - \Phi(A)|$ and (2) asymptotic

18

expansion of the difference $P_n(A) - \Phi(A)$, should be investigated for as wide a class of sets as possible.

In this connection our studies on the relation between weak convergence and uniform convergence of measures tell us that

$$(1.3.5) \qquad \sup_{A \in \mathcal{C}} |P_n(A) - \Phi(A)| \to 0$$

as $n \to \infty$, where $\mathcal{C}$ is the class of all convex subsets of R_k. Thus we may expect to find uniform bounds for the difference $P_n(A) - \Phi(A)$ for all convex sets. Further the class of convex sets is sufficiently wide to include all sets met with in practice such as rectangles, spheres, ellipsoids etc. Thus we may consider investigating the error, with regard to both types of problems mentioned above, for all convex sets. This investigation is carried out in the third part of the thesis.

Esseen [15] and Bergström [3] have studied the accuracy of the normal approximation for certain classes of sets in k-dimensional space. Bergström has shown that

$$|F_n(\underset{\sim}{x}) - \Phi(\underset{\sim}{x})| < C(k) \frac{\beta_3}{\sqrt{n}}$$

where $C(k)$ is a constant depending only on k and $\beta_3 = \sum_{j=1}^{k} E|Z_j^{(1)}|^3$. Esseen studies the error for sets A of the form

19

$$(1.3.5) \qquad A = [\, \underset{\sim}{x} : x_1^2 + \ldots + x_k^2 \leq r^2 \,]$$

for some r. His main result is the following:

$$(1.3.6) \qquad | P_n(A) - \Phi(A) | \leq C(k)\, \beta_4^{3/2}\, n^{-k/k+1}$$

where $C(k)$ depends only on k and the inequality (1.3.6) holds uniformly for all sets A of the form (1.3.5). The two authors follow widely different methods in obtaining their results. Esseen relies on characteristic functions as his main tool, while Bergström exploits a convolution identity and a result on the divergence $F(x) - G(x)$ of two distributions when the corresponding divergence $F * \Phi_\varepsilon - G * \Phi_\varepsilon$ is known, where Φ is a normal distribution with small variance ε. The method used in this thesis relies on characteristic functions, and is a variation of a method due to Esseen.

Apart from the results of Esseen and Bergström mentioned in the preceding paragraph no further results on the accuracy of normal approximation are known. In particular no theoretical work has been done on the problem of obtaining an asymptotic expansion for the difference $P_n(A) - \Phi(A)$. It is clear however that a formal expansion based on the expansion of characteristic functions analogous to (1.3.2) can be easily written down. Such a multivariate Edgeworth expansion has not been shown to be a valid asymptotic expansion. Wallace ([39], page 641) has drawn attention to this problem.

20

In this thesis we investigate this problem and obtain estimates of the remainders when the distribution function satisfies the condition

$$\limsup_{|\underset{\sim}{t}| \to \infty} |f(\underset{\sim}{t})| < 1 .$$

This analogue of Cramér's condition is satisfied when the d.f. has an absolutely continuous component. This condition is not satisfied for discrete distributions and once again the result then is not generally valid.

Among the discrete distributions which violate the Cramér's condition is the important class of lattice distributions in R_k. It may be recalled that for lattice distributions on R_1 Esseen obtained an asymptotic expansion by adding some discontinuous terms to the Edgeworth expansion. In R_k also we study this problem for lattice distribution and establish an asymptotic expansion for the difference $P_n(A) - \Phi(A)$ that is valid for all sets A.

The method followed here consists in deriving an asymptotic expansion for the magnitude of the jump of $F_n(x)$ at a discontinuity point. The passage to the probabilities $P_n(A) - \Phi(A)$ is then achieved by appealing to a generalization of the Euler - Maclaurin sum formula for functions of several variables. This generalization is given in a separate appendix (Appendix B) to this thesis.

21 It may be remarked that the contents of this part do not depend on Parts I and II and can be read independently.

1.4. Summary of the Thesis

The thesis consists of six chapters and two appendices. The six chapters are grouped into three parts. We have already indicated the problems considered in these parts and the motivation for their study. A brief summary of each chapter is given in the following paragraphs.

Part I (Chapters 2 and 3)

Let X be a complete separable metric space and μ_n a sequence of measures on X converging weakly to a measure μ on X. Then

1) $\int f \, d\mu_n \to \int f \, d\mu$ for each bounded continuous function on X, and

2) $\mu_n(A) \to \mu(A)$ for each continuity set A of μ.

In chapter 2 the uniformity of these convergences is investigated. It is shown that the convergence (1) above is uniform for all functions f belonging to any class $\mathcal{R}$ satisfying the conditions:

i) $\mathcal{R}$ is uniformly dominated by a continuous function $g(x)$,

ii) $\mathcal{R}$ is u.c.c. compact, and

iii) $\int g \, d\mu_n \to \int g \, d\mu \quad (< \infty)$

(Theorem 2.4.2).

1774

A similar result is obtained with regard to the convergence (2). The topology of uniform convergence on compacta, which turns out to be relevant to these questions, is discussed on section 2.2.

In chapter 3 applications of the general results are given to measures on (1) the k-dimensional Euclidean space R_k and (2) $C[0, 1]$ - the space of all continuous functions on the closed interval $[0, 1]$.

For measures on R_k uniformity of convergence is studied with regard to two classes of sets, the class (called $\mathcal{H}_m$) of all sets A which are intersections of at most m half spaces, and the class (called $\mathcal{C}$) of all measurable convex subsets of R_k. The main result is the following: if $\mathcal{C}$ ($\mathcal{H}_m$) is a continuity class for μ, then $\mu_n \Rightarrow \mu$ (converges weakly) if and only if $\mu_n(A) \to \mu(A)$ uniformly for all $A \in \mathcal{C}$ ($\mathcal{H}_m$). (Theorems 3.3.1 and 3.4.1). This chapter also contains a study of uniform convergence over the class $\mathcal{H}_m$ when it is not a continuity class for μ.

In the case of the space $C[0, 1]$, let P_n be a sequence of stochastic processes realised in $C[0, 1]$ converging weakly to the Brownian motion process (or a process that is absolutely continuous with respect to the Brownian motion process), say P. In this case the uniformity of convergence of $P_n(E) \to P(E)$ is established for a certain class of bands (Theorem 3.5.1). By a band we mean a subset $E \subset C[0, 1]$ of the form

23

$$E = [\, x(t);\quad \alpha(t) \le x(t) \le \beta(t);\ 0 \le t \le 1]$$

where $\alpha(t)$ and $\beta(t)$ are continuous functions.

Part II (Chapter 4)

Let $\xi_1, \xi_2, \ldots$ be a strictly stationary sequence of random elements in a complete separable metric space X. Let μ_n be the e.d.f. i.e. the measure on X with masses $\frac{1}{n}$ at the points $\xi_1, \ldots, \xi_n$. In this chapter we investigate the convergence properties of μ_n and its application to the strong law of large numbers for random functions.

Let $\mathcal{I}$ be the invariant σ-field associated with the sequence $\{\xi_n\}$. Then there exists a random measure on X, $\mu(A, w)$, such that, if $E|f(\xi_1)| < \infty$, then

$$E[\, f(\xi_1) / \mathcal{I}\,] = \int_X f(x)\, \mu(dx; w)$$

almost everywhere (Lemma 4.2.1). The random measure $\mu(A, w)$ arises as the limit of the sequence μ_n. In fact, $P[\, \mu_n \Rightarrow \mu] = 1$. (Theorem 4.3.1.). We then study convergence of the quantity

$$\eta_n = \sup_{f \in \mathcal{A}} \left| \int f\, d\mu_n - \int f\, d\mu \right|$$

where $\mathcal{A}$ is a class of continuous functions. It is shown that under certain conditions,

i) $P\left[\ \eta_n \to 0\ \right] = 1,$ and

ii) $E\,\eta_n^{1+\alpha} \to 0.$

(Theorem 4.3.2). This generalizes the earlier results of Fortet and Mourier, and also yields as an immediate consequence the strong law of large numbers for random elements in a separable Banach space.

Next we discuss the law of large numbers when $X = D\,[0, 1]$ - the space of all functions on the segment $[0, 1]$ with discontinuities, if any, only of the first kind. Let $\xi_n = \xi_n(t, w)$ be a strictly stationary sequence realised in $D[0, 1]$. Let

$$m\,(t, w) = E\,[\xi_1\,(t, w) / \mathcal{J}\,], \text{ and}$$

$$S_n\,(t, w) = [\,\xi_1(t, w) + \ldots + \xi_n\,(t, w)]/n.$$

The following result is proved: If $E\,\|\,\xi_1\,\|^{1+\alpha} < \infty$ for some $\alpha \geq 0$ (where $\|\xi_1\,(t, w)\,\| = \sup_t |\,\xi_1(t, w)\,|$), then

$$(a)\ P\left[\ \lim_{n \to \infty} \|\,S_n - m(t, w)\,\| = 0\ \right] = 1,$$

$$(b)\ \lim_{n \to \infty} E\,\|\,S_n - m\,(t, w)\,\|^{1+\alpha} = 0\,.$$

(Theorem 4.5.2).

In the last section of chapter 4 several generalizations of the classical Glivenko-Cantelli lemma to random vectors are obtained.

25

Let $X = R_k$, then it is shown that

$$\lim_{n \to \infty} \sup_{A \in \mathcal{H}_m} | \mu_n(A) - P[\xi_1 \in A / \mathcal{J}] | = 0$$

with probability one (Theorem 4.6.1). This result constitutes a considerable improvement over the earlier results of Wolfowitz, Fortet and Mourier, and Tucker.

Part III (Chapters 5 and 6)

Let $Z^{(1)}, Z^{(2)}, \ldots$ be a sequence of independent and identically distributed random vectors in R_k with mean zero and variance-covariance matrix = I. Let P_n be the probability distribution of the normed sum

$$\frac{Z^{(1)} + \ldots + Z^{(n)}}{\sqrt{n}},$$

and P the standard normal probability distribution. In Chapter 5 the accuracy of the normal approximation is investigated. A feature of this study is that the error of approximation $P_n(A) - P(A)$ is studied for a large class of sets, namely the class of all convex sets. The main result is the following inequality

$$\sup_{A \in \mathcal{C}} | P_n(A) - P(A) | \leq C(k)\, \beta_4^{3/2}\, \frac{(\log n)^{\alpha}}{\sqrt{n}}$$

where $\mathcal{C}$ is the class of all convex sets and $\alpha = \frac{(k-1)}{2(k+1)}$.

(Theorem 5.3.1).

An asymptotic expansion of the error $P_n(A) - P(A)$ is obtained for the case when the characteristic function $f(\underset{\sim}{t})$ of $Z^{(1)}$ satisfies the following analogue of Cramér's condition:

(C) $$\limsup_{|\underset{\sim}{t}| \to \infty} | f(\underset{\sim}{t}) | < 1$$

where $| \underset{\sim}{t} |^2 = t_1^2 + \ldots + t_k^2$. In this case it is shown that

$$P_n(A) = P(A) + \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} \int_A d\, \tilde{P}_j(-\Phi) + O\left(\frac{(\log n)^{\frac{1}{2}(k-1)}}{n^{\frac{1}{2}(s-2)}}\right)$$

uniformly for all convex sets A. (theorem 5.3.2).

The class of lattice distributions in R_k, for which the condition (C) does not hold, is discussed in Chapter 6. Let $Z^{(1)}$ be a lattice random vector and let $p_n(\underset{\sim}{z})$ denote the mass at a discontinuity point $\underset{\sim}{z}$ of the normed sum. If we suppose that $Z^{(1)}$ is concentrated in the lattice of vectors with integer coordinates and $\beta_s < \infty$, then

$$\sum_{\underset{\sim}{z}} | p_n(\underset{\sim}{z}) - q_n^{(s)}(\underset{\sim}{z}) | = 0 \left\{ \frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}} \right\} .$$

where

$$q_n^{(s)}(\underset{\sim}{x}) = n^{-\frac{1}{2}k} \sum_{j=0}^{s-3} n^{-\frac{1}{2}j} \tilde{P}_j(-\varphi)$$

(Theorem 6.3.2). This generalizes a result due to Prohorov for the Binomial case.

The asymptotic expansion of the error $P_n(A) - P(A)$ for the lattice case is deduced from the above theorem by utilizing a generalization of the Euler-Maclaurin sum formula to functions of several variables. The result in the simplest case is as follows:

If $F_n(\underset{\sim}{x}) = F_n(x_1, \ldots, x_k)$ denotes the d.f. of the normed sum, and $\beta_5 < \infty$, then

$$F_n(\underset{\sim}{x}) = \Phi(\underset{\sim}{x}) + \frac{1}{\sqrt{n}} \tilde{P}_1(-\Phi) - \frac{1}{\sqrt{n}} \sum_{j=1}^{k} S_1[x_j \sqrt{n} - na_j] \frac{\partial \Phi}{\partial x_j} + O\left(\frac{1}{n}\right)$$

where $S_1(t)$ for real t is a discontinuous function defined by $t - [t] - \frac{1}{2}$ (Theorem 6.4.2).

28

Appendices (A and B)

Appendix A is an exposition of such properties of convex sets, and of functions associated with convex sets, as are relevant to the thesis.

In appendix B a generalization of the Euler-Maclaurin sum formula to functions of several variables is given.

CHAPTER 2.

ON THE RELATION BETWEEN WEAK CONVERGENCE AND 'UNIFORM' CONVERGENCE OF MEASURES

2.1. Introduction The results of this Chapter are basic to the problems studied in Chapters 3 and 4. The Chapter consists of two parts. The first part begins with some preliminary material on the topological and measure theoretic terminology employed in the thesis, and a brief discussion of one or two results which are repeatedly used in subsequent chapters. Then weak convergence of measures is considered, and a brief survey of ~~the~~ certain important previously known results are given.

The second part is devoted to the study of the relation between weak convergence and uniform convergence over (1) a given class of functions, and (2) a given class of sets. These two problems, and the reasons for considering them, have already been described in the general introduction (Chapter 1). Satisfactory solutions to the two problems are obtained in this Chapter. The relevance of the work of LeCam in this connection is also pointed out.

2.2. Topological and measure-theoretic preliminaries Through out this Part, the symbol X will be used to denote an arbitary separable and complete metric space. Different symbols will be used to denote particular spaces, for instance, R_k always denotes the k-dimensional Euclidean space. By a function on X we always mean a real-valued function. C(X) is the collection of all bounded continuous functions on X. More generally, let

CALCUTTA CENTRAL LIBRARY
UNIVERSITY CENTRAL LIBRARY

$C_k(X)$ denote the collection of all continuous maps of X into R_k. Several topologies could be introduced in the space $C_k(X)$, and one of these will be frequently made use of in this thesis. We therefore discuss this particular topology at some length here.

Let $K \subset X$ be an arbitrary compact set and let $|u|$ for $u \in R_k$ denote the Euclidean norm there i.e. if $u = (u_1, u_2, \ldots, u_k)$ then

$$|u|^2 = \sum_{i=1}^{k} u_i^2$$

For an arbitrary $f_o \in C_k(X)$ and any $\varepsilon > 0$ let

$$N(f_o, K, \varepsilon) = [f;\ f \in C_k(X),\ \sup_{x \in K} |f(x) - f_o(x)| < \varepsilon]$$

Then it is not difficult to show that the class of all sets $\{N(f_o, K, \varepsilon)\}$ constitute a neighbourhood system for a topology. This topology is what is commonly known as the uniform convergence on compacta topology (u.c.c. topology). As its name implies, the u.c.c. topology can be described easily in terms of convergence. One may caution here that this topology is in general non-metrizable, and consequently a complete description in terms of convergence could be accomplished only with the aid of generalized sequences or Moore-Smith convergence. However, since only sequential convergence will be considered in this thesis we may confine ourselves to describing this convergence for sequences only.

Definition. A sequence of elements $f_n \in C_k(X)$ converges uniformly on compacta (u.c.c.) to an element $f_o \in C_k(X)$ if and only if for each compact set $K \subset X$,

$$\lim_{n \to \infty} \sup_{x \in K} |f_n(x) - f_o(x)| = 0.$$

A good deal is known about this convergence and the associated topology. (see Kelley [21] , chapter 7). Of particular importance to us to the structure of the sequentially compact subsets of $C_k(X)$. (It may be recalled that a subset $\mathcal{R} \subset C_k(X)$ is sequentially compact if and only if for each sequence $f_n \in \mathcal{R}$, there exists a subsequence f_{n_k} and an element $f_o \in \mathcal{R}$ such that f_{n_k} converges to f_o in u.c.c.). It is, however, much simpler to describe the nature of compact subsets of $C_k(X)$ by means of Ascoli's theorem (see Kelley [21]) as follows: $\mathcal{R} \subset C_k(X)$ is compact if and only if

(i) $\mathcal{R}$ is compact under pointwise convergence

(ii) the family $\mathcal{R}$ is equicontinuous on each compact set.

If the space X is locally compact then the condition (ii) stated above could be replaced by the condition (ii)' $\mathcal{R}$ is a equicontinuous family. Since X is also separable metric it is then not difficult to deduce from Ascloi's theorem that a compact family $\mathcal{R}$ is necessarily sequentially compact. In the general case, the following lemma points out certain sequentially compact families.

Lemma 2.2.2. Suppose $\mathcal{R} \subset C_k(X)$ satisfies the conditions.

(i) for each $x \in X$

$$\sup [\, | f(x) | , f \in \mathcal{R}] < \infty$$

(ii) there exists a constant M such that for $x, y \in X$ and $f \in$

$$| f(x) - f(y) | \leq M\, d(x, y)$$

where $d(x, y)$ is the metric of X.

Then the closure of the family $\mathcal{R}$ (in u.c.c. topology) is sequentially compact.

This lemma is well known and its proof will be omitted.

2.2.1. Measures on X. Let $\mathcal{B}$ denote the class of all Borel subsets of X. It might be recalled that $\mathcal{B}$ is the σ-field generated by the class of all closed subsets of X. All subsets of X considered here are Borel sets and this will not be explicitly mentioned henceforth. A measure on X is a nonnegative, countably additive, totally finite set function on $\mathcal{B}$. We sometimes use the term distribution to denote a normalised measure i.e. a measure μ for which $\mu(X) = 1$. A measure is continuous (= non-atomic) if the μ-measure of each single point set is zero.

If μ is a measure on X and A is a Borel set, then μ_A denotes the restriction of μ to the set A, i.e.

$$\mu_A(E) = \mu(A \cap E) \quad \text{for each} \quad E \in \mathcal{B}.$$

If $f(x)$ is a measurable map of X into another metric space Y, then μf^{-1} denotes the measure on the Borel sets of Y defined by the relation

$$\mu f^{-1}(E) = \mu[f^{-1}(E)]$$

for each Borel subset E of Y.

In the following paragraphs we give a brief description of a concept of convergence for measures which has become important in probability theory.

2.3. Weak convergence of measures. For distributions on the real line this concept was introduced and studied long ago, notably by Montel, Helly Levy and others. A sequence F_n of distribution functions on the real line is said to converges weakly to F if $F_n(x) \to F(x)$ at every continuity point of the latter. Though this definition is not amenable to gene-

generalisation, there are equivalent forms of this convergence which may be used for this purpose. For measures on X we introduce the following definition.

Definition. A sequence of measures μ_n $(n = 1, 2, ..)$ converges weakly to a measure μ (in symbols, $(\mu_n \Rightarrow \mu)$ if and only if

$$\lim_{n \Rightarrow \infty} \int f d\mu_n = \int f d\mu$$

for each $f \in C(X)$.

A study of this concept of weak convergence was begun by Alexandroff in his memoirs in Mat. Sbornik ([1]). He studied this in the very general framework of measures on an arbitrary topological space, and proved some important theorems. This study was continued by Prohorov ([29]), LeCam ([25]), and Varadarajan ([35]). While Prohorov restricts himself to distributions on a separable and complete metric space, the last two authors carry out the study in the framework of measures on a general topological space. However, as mentioned earlier, we restrict ourselves to the level of generality of a separable and complete metric space, even though many of the results of this thesis could be proved for more general spaces. For the sake of continuity and ease of reference, a brief survey of the main results of this theory will be given below. For all these results reference is made to (Prohorov [29], Varadarajan [35], [36]).

Let $d(x, y)$ denote the metric in X and let $U_d(X)$ denote the family of all bounded functions on X which are uniformly continuous with respect to the metric d. If should be noted that the class $U_d(X)$

varies with d and $U_d(X) \subset C(X)$.

If μ is a measure on X, then a set A is a continuity set for μ if the μ-measure of the boundary of A is zero. A function f(x) on X is said to be almost everywhere continuous with respect to μ if

$$\mu^*(X - C_f) = 0$$

where C_f is the set of all points of continuity of f, and μ^* denotes the outer measure generated by μ.

Theorem 2.3.1. The following statements are equivalent:

(1) $\mu_n \Rightarrow \mu$

(2) $\int f d\mu_n \Rightarrow \int f d\mu$ for each $f \in U_d(X)$

(3) $\int f d\mu_n \Rightarrow \int f d\mu$ for each bounded function f which is almost everywhere continuous with respect to μ.

(4) $\mu_n(A) \Rightarrow \mu(A)$ for each continuity set A of μ.

(5) $\mu_n f^{-1} \Rightarrow \mu f^{-1}$ for each $f \in C(X)$.

For the proof and other details the reader is referred to [29]. Note that the condition (2) of the above Theorem requires convergence only over the class $U_d(X)$. By suitably altering the metric one could make the class $U_d(X)$ smaller. In fact, since X is a separable and complete metric space an equivalent metric D(x, y) could be found which makes $U_D(X)$ a separable space under the sup norm (see Varadarajan [36], p 18). We have as a consequence the following important theorem due to Varadarajan ([37], p 14).

Theorem 2.3.2. There exists a fixed sequence $f_1, f_2, \ldots$ of functions $\in C(X)$ such that the following two statements are equivalent:

(i) $\mu_n \Rightarrow \mu$

(ii) for each i, $\int f_i d\mu_n \Rightarrow \int f_i d\mu$ as $n \Rightarrow \infty$.

One of the consequences of Theorem 2.3.2. is that the space of measures $\mathcal{M}(X)$ with the weak topology is metrizable. It could also be shown that $\mathcal{M}(X)$ could be metrized as a separable and complete metric space. (Varadarajan [36]). This fact has been established earlier by Prohorov [29], who used a different method.

One of the important results on the structure of the weak topology is the characterization of compact subsets of $\mathcal{M}(X)$. A complete solution to this problem is contained in Prohorov [29]. He proves the following.

THEOREM 2.3.3. Suppose a subset $\mathcal{N} \subset \mathcal{M}(X)$ satisfies the following two conditions.

(i) $\sup [\, \mu(X),\ \mu \in \mathcal{N} \,] < \infty$

(ii) for each $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset X$ such that $\mu(X - K_\varepsilon) \leq \varepsilon$ for all $\mu \in \mathcal{N}$.

Then the family $\mathcal{N}$ is conditionally compact (equivalently, the closure of the family $\mathcal{N}$ is compact).

Theorem 2.3.3. was utilized by Prohorov to obtain limit theorems for random functions. The theorem will play a crucial role in the establishment of the main theorems of this chapter.

We conclude this section with a lemma which is an immediate consequence of the above results and which will be required later.

Lemma 2.3.1. Let μ_n be a sequence of measures on X converging weakly to μ and let E be a closed continuity set of μ. Then the sequence $(\mu_n)_E$, considered as measures on the metric space E. (relativized topology), converges weakly to the measure μ_E.

Proof. Let f be an arbitrary bounded continuous function on E i.e. $f \in C(E)$. Define an associated function $\hat{f}$ on X as follows

$$\hat{f}(x) = \begin{cases} f(x) & x \in E \\ 0 & x \notin E \end{cases}$$

Then, since E is a continuity set of the measure μ, it follows without difficulty that f is almost everywhere continuous with respect to μ. Consequently, by Theorem 2.2,

$$\int \hat{f}\, d\mu_n \Rightarrow \int \hat{f}\, d\mu.$$

Hence, from the definition of $\hat{f}$,

$$\int_E f\, d\mu_n \Rightarrow \int_E f\, d\mu$$

i.e. for each $f \in C(E)$, $\int f d(\mu_n)_E \Rightarrow \int f\, d\mu_E$. This completes the proof.

2.4. Uniform convergence over a class of functions. In this section the relation between weak convergence ($\mu_n \Rightarrow \mu$) and uniform convergence of $\int f d(\mu_n - \mu)$ over a class of functions f will be studied. The main tool used here is the compactness criterion of Prohorov (theorem 2.3.3.). The two main theorems of this section are easy consequences of this criterion. The results of this section will be used later in connection with the law of large numbers for random functions.

THEOREM. 2.4.1. Let $\mathcal{A}$ be any given family of continuous functions on X possessing the following two properties: (i) there exists a constant M such that for all $x \in X$ and $f \in \mathcal{A}$, $|f(x)| \leq M$. (ii) $\mathcal{A}$ is compact (u.c.c. topology).

If μ_n, μ $(n = 1, 2, \ldots,)$ are measures on X then $\mu_n \Rightarrow \mu$ if and only if

$$(2.4.1) \qquad \lim_{n \to \infty} \sup_{f \in \mathcal{R}} \left| \int f d\mu_n - \int f d\mu \right| = 0.$$

Proof. The 'if' part is trivial. To prove the necessity of (2.4.1), let us write $\alpha_n(f) = \left| \int f d\mu_n - \int f d\mu \right|$. Take any $f_o \in \mathcal{R}$ and consider a neighbourhood of f_o (on u.c.c. topology) defined by

$$N(f_o : K : \varepsilon) = \left[f : \sup_{x \in K} |f(x) - f_o(x)| \leq \varepsilon \right]$$

where K is some compact subset of X. Then it is easily deduced that for each $f \in N(f_o : K : \varepsilon)$

$$\alpha_n(f) \leq \left| \int f_o d\mu_n - \int f_o d\mu \right| + \int_K |f - f_o| \, d(\mu_n + \mu) + \int_{X-K} |f - f_o| \, d(\mu_n + \mu)$$

$$(2.4.2) \qquad \alpha_n(f) \leq \alpha_n(f_o) + \varepsilon(\mu_n + \mu)(X) + 2M(\mu_n + \mu)(X - K)$$

Since $\mu_n \Rightarrow \mu$, it follows from the compactness criterion (Theorem 2.3.3) that there exists a compact set K_ε such that

$$\mu_n(X - K_\varepsilon) < \varepsilon, \quad \mu(X - K_\varepsilon) < \varepsilon$$

for all n. It thus follows from (2.4.2) that

$$(2.4.3) \qquad \limsup_{n \to \infty} \sup \left[\alpha_n(f) : f \in N(f_o ; K_\varepsilon ; \varepsilon) \cap \mathcal{R} \right] \leq 2\varepsilon(\mu(X) + 2M).$$

Now, it is clear that the $N(f_o ; K_\varepsilon ; \varepsilon)$, for f_o ranging over $\mathcal{R}$, constitute an open covering of $\mathcal{R}$. Since $\mathcal{R}$ is compact there is a finite subcovering, i.e. there exist elements $f_1, f_2, \ldots f_k \in \mathcal{R}$ such that

$$(2.4.4) \qquad \bigcup_{i=1}^{k} N(f_i; K_\varepsilon: \varepsilon) \supset \mathcal{R}$$

Combining (2.3.3) and (2.3.4), it follows that

$$(2.4.5) \qquad \limsup_{n \to \infty} \sup [\alpha_n(f): f \in \mathcal{R}] \leq 2\varepsilon(\mu(X) + 2M).$$

Since ε is arbitrary, (2.4.1) from (2.4.5). This completes the proof.

In the above theorem the class $\mathcal{R}$ was assumed to be uniformly bounded. Our next result removes this restriction.

Theorem 2.4.2. Let $\mathcal{R}$ be a family of continuous functions satisfying the following two conditions: (i) there exists a continuous function $g(x)$ on X such that $|f(x)| \leq g(x)$ for all $x \in X$ and all $f \in \mathcal{R}$. (ii) $\mathcal{R}$ is compact (u.c.c. topology).

Suppose that $(\mu_n, \mu, n = 1, 2, \ldots,)$ is a sequence of measures on X such that (a) $\mu_n \Rightarrow \mu$, (b) $\int g d\mu < \infty$, and (c) $\int g d\mu_n \to \int g d\mu$. Then

$$(2.4.6) \qquad \lim_{n \to \infty} \sup_{f \in \mathcal{R}} \left| \int f d\mu_n - \int f d\mu \right| = 0.$$

Proof. Let $E_N = [x : g(x) \leq N]$. Since $g(x)$ is continuous, it follows that E_N is closed and that its boundary is $[x: g(x) = N]$. We suppose that N is so chosen that E_N becomes a continuity set for μ. (i.e. $\mu [x: g(x) = N] = 0$. Since E_N is a closed subset of X, it is also a separable and complete metric space and it is easily verified that the family of functions $\mathcal{R}$ restricted to the set E_N satisfy both the conditions (i) and (ii) of theorem 2.4.1. It follows from Lemma 2.2.1 that $(\mu_n)_E \Rightarrow \mu_E$. Thus all the conditions of theorem 2.4.1 are fulfilled, and we have

$$(2.4.7)\qquad \lim_{n \to \infty} \sup_{f \in \mathcal{A}} \left| \int_{E_N} f \, d\mu_n - \int_{E_N} f d\mu \right| = 0$$

for each N for which E_N is a continuity set.

Now the fact that E_N is a continuity set of μ implies that the function $\hat{g}(x)$ defined by

$$\hat{g}(x) = \begin{cases} g(x) & x \in E_N \\ 0 & \text{otherwise} \end{cases}$$

is almost everywhere continuous with respect to μ and hence

$$\lim_{n \to \infty} \int_{E_N} g(x) d\mu_n \to \int_{E_N} g(x) d\mu .$$

The assumption (b) then implies that

$$(2.4.8)\qquad \lim_{n \to \infty} \int_{X-E_N} g(x) d\mu_n = \int_{X-E_N} g(x) d\mu .$$

Finally, since $|f(x)| \leq g(x)$ for $f \in \mathcal{A}$, it follows that

$$(2.4.9)\qquad \sup_{f \in \mathcal{A}} \left| \int f d\mu_n - \int f d\mu \right| \leq \sup_{f \in \mathcal{A}} \left| \int_{E_N} f d\mu_n - \int f d\mu \right| + \int_{X-E_N} g d\mu_n + \int_{X-E_N} g d\mu .$$

Combining (2.4.7), (2.4.8) and (2.4.9) one obtains

$$(2.4.10)\qquad \overline{\lim_{n \to \infty}} \sup_{f \in \mathcal{A}} \left| \int f d\mu_n - \int f d\mu \right| \leq 2 \int_{X-E_N} g d\mu .$$

If in (2.4.10) N is allowed to tend to infinity through values for which E_N is a continuity set, then one obtains (2.4.6).

This completes the proof of the theorem.

Theorems 2.4.1 and 2.4.2 provide answers to the first problem raised in the introduction. The solutions obtained here are quite general from the point of view of our applications.

Next, we shall deduce the following result due to Prohorov ([29], p. 186) as a corollary to Theorem 2.4.1.

Theorem 2.4.3. Let μ_n, (n = 1, 3, ...) be a sequence of measures on X converging weakly to a measure μ. Let $\xi_n(x)$ be a sequence of continuous maps of X into another metric space Y, converging uniformly on compacta to $\xi(x)$. Then $\mu_n \xi_n^{-1} \Rightarrow \mu \xi^{-1}$.

Proof. Let g(y) be an arbitrary bounded uniformly continuous function on Y. Let $h_n(x) = g[\xi_n(x)]$ and $h(x) = g[\xi(x)]$. Since $\xi_n(x)$ converges (u.c.c.) to $\xi(x)$, it follows that the sequence of functions $h_n(x)$ converges uniformly on compacta to h(x). Since these functions are uniformly bounded, it follows from theorem 2.4.1 that

$$\lim_{n \to \infty} [\int h_n(x)d\mu_n - \int h_n(x)d\mu] = 0$$

Also, from the bounded convergence theorem it is clear that

$$\int h_n(x)d\mu \to \int h(x)d\mu.$$

Thus for each bounded and uniformly continuous function g(y) on Y, one has

$$\int g[\xi_n(x)]d\mu_n \to \int g[\xi(x)]d\mu$$

or equivalently

$$\int g(y)d\mu_n \xi_n^{-1} \to \int g(y)d\mu \xi^{-1}.$$

Consequently, by theorem 2.3.1, we have that $\mu_n \xi_n^{-1} \Rightarrow \mu \xi^{-1}$. This completes the proof.

Remarks. Theorem 2.4.1. above is in fact equivalent to the compactness criterion of Prohorov. This follows from a result on a characterization(due to LeCam [25], p 217) of the so called 'tight' functionals. Indeed, one might say that Theorem 2.4.1 is essentially contained in the result of LeCam although LeCam does not state it in the form given here. LeCam's approach consists in viewing the measure μ as a linear functional, $\mu(f) = \int f d\mu$. Let S(X) denote the unit sphere of C(X) and endow S(X) with the u.c.c. topology. Then LeCam proves that the following two are equivalent:

(i) the family $\mu(f)$, $\mu \in \mathcal{T} \subset \mathcal{M}(X)$ of functionals on S(X) are equicontinuous at the origin (with respect to the uniformity induced by the vector structure of C(X)),

(ii) the family $\mathcal{T}$ is uniformly 'tight' i.e. given any $\varepsilon > 0$, there exists a compact set K_ε such that $\mu(X - K_\varepsilon) \leq \varepsilon$ for all $\mu \in \mathcal{T}$.

Thus it is clear that Theorem 2.4.1. could be deduced from the above result, the compactness criterion (theorem 2.3.3) and the Ascoli theorem. (see Kelley, Chapter 7). However, it appears to us that the approach presented in this thesis is particularly simple.

2.5. Uniform convergence over a class of sets. In this section the relation between weak convergence and uniform convergence over a class of sets is investigated. The main theorem of this section isolates classes of sets $\mathcal{B}$ with the following property: if $\mathcal{B}$ is a continuity class for μ, and if μ_n converges weakly to μ, then $\mu_n(A)$ converges to $\mu(A)$ uniformly over $\mathcal{B}$. The method employed here consists in reducing the problem essentially to finite dimensional Euclidean spaces with the

help of a theorem of Prohorov (Theorem 2.4.5).

In this section a class $\mathcal{B}$ is said to be a continuity class for a measure μ if each set $A \in \mathcal{B}$ is a continuity set for μ.

We begin with a discussion of the problem in a particular case. Here $X = R_k$ - the k dimensional Euclidean space. We start with a fixed coordinate system in R_k. If μ is a measure on R_k, then the projections of μ on the axes are called the marginal measures or the marginal components of μ. By a rectangle $E \subset R_k$ is meant a set of the form $E = \{ (x_1, \ldots x_k);\ x_i \in I_i;\ i = 1, 2, \ldots k \}$ where I_i are closed, open or semi-open intervals. We use the symbol $\mathcal{R}_k$ to denote the collection all rectangles of R_k.

Lemma 2.5.1. Let μ, μ_n $(n = 1, 2, \ldots)$ be a sequence of measures on R_k. Suppose that the class of sets $\mathcal{R}_k$ is a continuity class for the measure μ. Then $\mu_n \Rightarrow \mu$ if and only if

$$\lim_{n \to \infty} \varrho(\mu_n, \mu) = 0 \tag{2.5.1}$$

where the distance $\varrho(\mu, \nu)$ is defined by

$$\varrho(\mu, \nu) = \sup [\ |\mu(E) - \nu(E)|,\ E \in \mathcal{R}_k\ ] \tag{2.5.2}$$

Proof. It is obvious that ϱ-convergence implies weak convergence. To prove the converse, let $\mathcal{A}$ denote the class of sets of the form $[(x_1, x_2, \ldots x_k);\ x_1 < a_1, \ldots x_k < a_k]$. It is then clear that $\varrho(\mu_n, \mu) \to 0$ if and only if

$$\sup_{A \in \mathcal{A}} |\ \mu_n(A) - \mu(A)\ | \to 0, \text{ as } n \to \infty. \tag{2.5.3}$$

Now, the assumption that $\mathcal{R}_k$ is a continuity class is equivalent to the statement that the marginal measures of μ are all continuous. Consequently, constants a_{ij} ($j = 0, 1, 2, \ldots r_i$, and $i = 1, 2, \ldots k$) can be found such that

$$(2.5.4) \qquad \begin{aligned} &-\infty = a_{i0} < a_{i1} < \ldots < a_{ir_i} = +\infty, \text{ and} \\ &\mu\,[X : a_{ij} \leq x_i \leq a_{ij+1}] \leq \varepsilon \end{aligned}$$

for all i and j, where x_i denotes the ith coordinate of the vector $x \in R_k$. Let $\mathcal{A}_\varepsilon$ be the class of all sets $\varepsilon\mathcal{A}$ with vertices at the points $(a_{1j_1}, a_{2j_2}, \ldots a_{kj_k})$. Then the class $\mathcal{A}_\varepsilon$ is finite and since the marginal measures of μ are continuous it follows that each $A \in \mathcal{A}_\varepsilon$ has μ-null boundary. Therefore $\mu_n \Rightarrow \mu$ implies that

$$(2.5.5) \qquad \sup_{A \in \mathcal{A}_\varepsilon} | \mu_n(A) - \mu(A) | \to 0, \quad \text{as} \quad n \to \infty$$

The next step of the proof consists in establishing the inequality

$$(2.5.6) \qquad \sup_{A \in \mathcal{A}} | \mu_n(A) - \mu(A) | \leq \sup_{A \in \mathcal{A}_\varepsilon} | \mu_n(A) - \mu(A) | + k\,\varepsilon$$

Suppose for a moment that the inequality (2.5.6) is true, then it follows from (2.5.5) that

$$(2.5.7) \qquad \limsup_{n \to \infty} \sup_{A \in \mathcal{A}} | \mu_n(A) - \mu(A) | \leq k\,\varepsilon$$

Since ε is arbitrary, the truth of (2.5.3) and hence that of the theorem would then follow from (2.5.7). It thus remains to prove (2.5.6).

Let $A \in \mathcal{A}$ be arbitrary. Then there is a vector $(b_1, b_2, \ldots b_k)$ such that $A = [x : x_i < b_i \text{ for each } i]$. Let a_{sj_s} be defined by the inequality

$$a_{sj_s} \leq b_s < a_{sj_s+1}$$

Let

$$A_1 = [x: x_s < a_{sj_s} \text{ for each } s]$$

$$A_2 = [x: x_s < a_{sj_s+1} \text{ for each } s].$$

Then it is clear that $A_1 \subset A \subset A_2$. Further,

$$\mu(A_2 - A_1) \leq \sum_{s=1}^{k} \mu[x: a_{sj_s} \leq x_s \leq a_{sj_s+1}]$$

$$\leq k \varepsilon .$$

The last inequality is a consequence of the choice made in (2.5.4). Since $A_1 \subset A \subset A_2$ it is not difficult to verify the inequality

$$|\mu_n(A) - \mu(A)| \leq \text{Max} \begin{Bmatrix} |\mu_n(A_1) - \mu(A_1)| \\ |\mu_n(A_2) - \mu(A_2)| \end{Bmatrix}$$

$$+ \mu(A_2 - A_1)$$

i.e.

$$|\mu_n(A) - \mu(A)| \leq \sup_{A \in \mathcal{A}_\varepsilon} |\mu_n(A) - \mu(A)| + k \varepsilon.$$

This proves the inequality (2.5.6) and thus completes the proof of the lemma.

The main theorem of this section may now be stated as follows.

Theorem 2.5.1. Let $\mathcal{A}$ be an arbitrary family of continuous maps from X into R_k which is sequentially compact (u.c.c. topology). Let $\mathcal{B}$ be the class of sets defined by $\mathcal{B} = [A: A = f^{-1}(E), E \in \mathcal{R}_k$ and $f \in \mathcal{A}]$. Suppose μ is a measure on X for which the class $\mathcal{B}$ is a continuity class or, equivalently, μf^{-1} has continuous marginal components for each $f \in \mathcal{A}$. Then $\mu_n \Rightarrow \mu$ implies that $\varrho(\mu_n f^{-1}, \mu f^{-1})$ converges to zero uniformly for $f \in \mathcal{A}$; equivalently,

$$\lim_{n \to \infty} \sup_{A \in \mathcal{B}} |\mu_n(A) - \mu(A)| = 0.$$

Proof. Suppose the contrary. Then there exists a $\delta > 0$ and a sequence of functions $f_n \in \mathcal{A}$ such that, for all n

$$\varrho(\mu_n f_n^{-1}, \mu f_n^{-1}) > \delta > 0. \tag{2.5.8}$$

Since $\mathcal{A}$ is sequentially compact there is a subsequence f_{n_k}, and a $f_o \in \mathcal{A}$ such that $f_{n_k} \to f_o$ uniformly on compact sets. In view of Theorem 2.4.5 it follows that

(i) $\mu_{n_k} f_{n_k}^{-1} \Rightarrow \mu f_o^{-1}$

(ii) $\mu f_{n_k}^{-1} \Rightarrow \mu f_o^{-1}$

Lemma 2.5.1 then implies that

$$\lim_{k \to \infty} \varrho(\mu_{n_k} f_{n_k}^{-1}, \mu f_o^{-1}) = 0 \tag{2.5.9}$$

$$\lim_{k \to \infty} \rho(\mu f_{n_k}^{-1}, \mu f_o^{-1}) = 0 \tag{2.5.10}$$

It follows from the triangle inequality that

$$\rho(\mu_{n_k} f_{n_k}^{-1}, \mu f_{n_k}^{-1}) \leq \rho(\mu_{n_k} f_{n_k}^{-1}, \mu f_o^{-1}) + \rho(\mu f_o^{-1}, \mu f_{n_k}^{-1}). \tag{2.5.11}$$

Taking the limit as $k \to \infty$ in (2.5.11), and utilizing (2.5.9) and (2.5.10), one obtains

$$\lim_{n \to \infty} \rho(\mu_{n_k} f_{n_k}^{-1}, \mu f_{n_k}^{-1}) = 0$$

which contradicts (2.5.8). This completes the proof of the theorem.

Remarks. Theorem 2.5.1 is a generalization of a theorem of Polya. The theorem of Polya referred to states that if $F(x)$ is a continuous distribution and $F_n(x)$ is a sequence of distribution functions (on the real line R) converging weakly to F, then

$$\lim_{n \to \infty} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)| = 0 .$$

This result is a special case of theorem 2.5.1 with X = the real line, and $\mathcal{A}$ the set containing the single function $f(x) = x$.

The above theorems (2.5.1) consider uniform convergence only over a continuity class. This is natural, since weak convergence in general implies only convergence for a continuity set. In the extreme case of X being discrete every set is a continuity set for every measure and in this case, as is well known, weak convergence is equivalent to uniform convergence over the class of all Borel sets.

CHAPTER 3

APPLICATIONS TO FINITE DIMENSIONAL EUCLIDEAN SPACES AND THE SPACE C[0,1].

3.1. Introduction. In this chapter we consider applications of Theorem 2.5.1 to two particular spaces. These are (a) the Euclidean space of finite dimension, and (b) the space C[0, 1] consisting of all continuous functions on the interval [0, 1]. We naturally study uniform convergence only over certain classes of sets peculiar to these spaces.

For measures on Euclidean spaces, uniform convergence over two different classes of sets is considered. They are (1) $\mathcal{H}_m$ = the class of all sets which are intersections of at most m half-spaces, and (2) $\mathcal{C}$ = the class of all measurable convex sets. It is shown here that weak convergence together with the continuity of the class ($\mathcal{H}_m$ or $\mathcal{C}$) ensures uniform convergence over that class. (A detailed study of the relevant properties of convex sets needed here is given in an Appendix to the thesis). Here we also obtain some sufficient conditions when the class $\mathcal{H}_m$ may cease to be a continuity class. This last investigation is motiveted by an application made in Chapter 4 in generalizing the classical Glivenko-Cantelli lemma.

In the case of the space C[0, 1] we consider a sequence of distributions converging weakly to Brownian motion process. The class of sets considered here is a certain class of sets of the form

$$[x(t) : \alpha(t) \leq x(t) \leq \beta(t), \text{ for } 0 \leq t \leq 1]$$

where $\alpha(t)$ and $\beta(t)$ satisfy certain conditions.

3.2. Some lemmas on measures in R_k. In this section a decomposition theorem for measures in R_k generalizing the usual decomposition into atomic and non-atomic components will be obtained. This decomposition theorem will be used in connection with measures for which the class $\mathcal{H}_m$ is not a continuity class.

The following notation and definitions will be used. A function $l(x)$ with or without suffixes always denotes a linear function on R_k. A subset of the form = $[x : l(x) = p]$ is called a hyperplane, while a half space is a set of the form = $[x : l(x) \leq p]$. The symbol $\mathcal{H}_m$ will denote the collection of all sets which are intersections of at most m half-epaces. In other words a set $A \in \mathcal{H}_m$ is of the form

$$A = [x : l_1(x) < p_1, \ldots, l_r(x) < p_r; \ r \leq m]$$

By a translate of a set E is meant any one of the sets $A+z$, $z \in R_k$, where $A+z = [(x + z); x \in A]$.

Definition 3.2.1. A measure λ on R_k is said to be r-dimensional $(0 \leq r \leq k)$ provided (i) there exists an r-dimensional subspace or a translate of such a subspace, say E, such that $\lambda(R_k - E) = 0$ and (ii) $\lambda(A) = 0$ for every set A which is a translate of a subspace of dimension less than r.

According to the above definition, a 0-dimensional measure is degenerate. We shall now obtain a decomposition of a measure into components of various different dimensions.

Lemma 3.2.1. Let λ be any measure and $\{A_\alpha\}$ be a family of sets such that $\lambda(A_\alpha \cap A_\beta) = 0$ for $\alpha \neq \beta$. Then $\lambda(A_\alpha) = 0$ for all but a countable number of sets A_α.

This result is well known and its proof will be omitted.

Lemma 3.2.2. For $0 \leq r \leq k-1$, let $\mathcal{V}_r$ be the collection of all sets A which are translates of r-dimensional subspaces of R_k. Then for any two sets $A, B \in \mathcal{V}_r$ $(1 \leq r \leq k-1)$ either (i) $A = B$ or (ii) $A \cap B = 0$ or (iii) $A \cap B \in \mathcal{V}_s$ for some s, $0 \leq s < r$.

Proof. From the definition of the class $\mathcal{V}_r$ it follows that there exist r-dimensional subspaces V_1 and V_2 and points z_1 and $z_2 \in R_k$ such that

$$A = V_1 + z_1; \qquad B = V_2 + z_2. \tag{3.2.1}$$

Now suppose that $A \cap B \neq 0$. If $z_o \in A \cap B$ then obviously $z_1 - z_o \in V_1$ and $z_2 - z_o \in V_2$ and therefore $V_1 + (z_1 - z_o) = V_1$ and $V_2 + (z_2 - z_o) = V_2$. In view of this, (3.2.1) may be rewritten as

$$A = V_1 + z_o; \qquad B = V_2 + z_o.$$

From this it follows at once that $A \cap B = V_1 \cap V_2 + z_o$. Now there are two possibilities. Either $\dim (V_1 \cap V_2) = r$, in which case $V_1 = V_2$ and therefore $A = B$; or $\dim (V_1 \cap V_2) < r$, in which case $A \cap B$ is a translate of an s-dimensional subspace for some $s < r$ and consequently belongs to $\mathcal{V}_s$. The proof of the lemma is thus complete.

Lemma 3.2.3. Let λ be any measure on R_k. Then there exists measures $\lambda_r^{(i)}$ $(r = 1, 2, \ldots k-1;\ i = 0, 1, 2, \ldots,)$ and a measure $\lambda_k^{(o)}$ all mutually orthogonal, such that

$$\text{(i)}\quad \lambda = \sum_{r=0}^{k-1} \sum_{i=0}^{\infty} \lambda_r^{(i)} + \lambda_k^{(o)}$$

(ii) for each r, $o \leq r \leq k-1$, and any i, $\lambda_r^{(i)}$ is a r-dimensional measure.

(iii) $\lambda_k^{(o)}$ is a k-dimensional measure.

Proof. Consider the class $\mathcal{V}_o$. It consists of all single point sets. It follows therefore that λ can be written in the form

$$\lambda = \sum_{i=0}^{\infty} \lambda_o^{(i)} + \mu_1$$

where $\lambda_o^{(i)}$ are the atoms of λ and the μ_1-measure of each set $A \in \mathcal{V}_o$ is zero. Now consider the measure μ_1 and the class $\mathcal{V}_1$. For any two distinct sets A and $B \in \mathcal{V}_1$, either $A \cap B = 0$ or $A \cap B \in \mathcal{V}_o$ and consequently $\mu_1(A \cap B) = 0$ always. By Lemma 3.2.1. there exist a countable number of sets $V_{1i} \in \mathcal{V}_1$ $(i = 0, 1, 2, \ldots)$ such that $\mu_1(A) = 0$ if $A \in \mathcal{V}_1$ and $A \neq A_i$ for some i. Define $\lambda_1^{(i)} = (\mu_1)_{A_i}$. Then it is clear that μ_1 can be written as

$$\mu_1 = \sum_{i=0}^{\infty} \lambda_1^{(i)} + \mu_2$$

It is obvious from the construction that all the measures μ_2 and $\lambda_1^{(i)}$ are all mutually orthogonal. Also for each $A \in \mathcal{V}_1$, $\mu_2(A) = 0$. Further, it is clear that $\lambda_1^{(i)}(A) = 0$ for $A \in \mathcal{V}_o$ $(i = 1, 2, \ldots)$ i.e. the measures $\lambda_1^{(i)}$ are all 1-dimensional measures. Thus λ may be written as

$$\lambda = \sum_{i=0}^{\infty} \lambda_o^{(i)} + \sum_{i=0}^{\infty} \lambda_1^{(i)} + \mu_2$$

where $\mu_2(A) = 0$ for $A \in \mathcal{V}_1$ and $\lambda_o^{(i)}$, $\lambda_1^{(i)}$ are respectively zero and 1-dimensional. It is then clear that the same argument applies to the measure

μ_2 and the class $\mathcal{V}_2$. Repeating the argument in this way one obtains that

$$\lambda = \sum_{r=0}^{k-1} \sum_{i=0}^{\infty} \lambda_r^{(i)} + \mu_k \tag{3.2.2}$$

where $\lambda_r^{(i)}$ are r-dimensional measures and $\mu_k(A) = 0$ for $A \in \mathcal{V}_{k-1}$. Then clearly μ_k is k-dimensional and thus the truth of the lemma follows from (3.2.2) with $\lambda_k^{(o)} = \mu_k$. The proof is thus complete.

3.3. Uniform convergence over the class $\mathcal{H}_m$. In this section Theorem 2.5.1. of Chapter 2, section 4, will applied to discuss the relation between weak convergence and uniform convergence over the class $\mathcal{H}_m$. The case when the class $\mathcal{H}_m$ is not a continuity class will also be considered.

Theorem 3.3.1. Let μ be a measure on R_k such that the μ-measure of each hyperplane is zero (or equivalently, the class $\mathcal{H}_m$ is a continuity class). Then $\mu_n \Rightarrow \mu$ if and only if

$$\lim_{n \to \infty} \sup_{A \in \mathcal{H}_m} | \mu_n(A) - \mu(A) | = 0 \tag{3.3.1}$$

for each m.

Proof. It is trivial to see that (3.3.1) implies weak convergence. To prove the necessity of (3.3.1), consider the class $\mathcal{A}_m$ of all linear maps of norm one of R_k into R_m. It is then obvious that $\mathcal{A}_m$ is sequentially compact. Also the class $\mathcal{B}_m = [A; A = f^{-1}(E), f \in \mathcal{A}_m$ and E a rectangle in $R_m]$ is a continuity class for μ. Thus Theorem 2.5.1. applies and yields

$$(3.3.2) \qquad \lim_{n \to \infty} \sup_{A \in \mathcal{B}_m} | \mu_n(A) - \mu(A) | = 0$$

for each m. Since $\mathcal{H}_m = \bigcup_{i \leq m} \mathcal{B}_{i,m}$, (3.3.1) follows from (3.3.2). The proof is complete.

An extension of the above result to the case where the family $\mathcal{H}_m$ is not necessarily a continuity class will now be obtained. Lemma 3.2.3 on decomposition of measures will be used for this purpose.

Lemma 3.3.1. Let E be a translate of a r-dimensional subspace, and μ a r-dimensional measure concentrated in E, i.e. $\mu(R_k - E) = 0$. If μ_n is a sequence of measures for which (a) $\mu_n \Rightarrow \mu$ and (b) $\mu_n(E) \to \mu(E)$ then (3.3.1) holds.

Proof. Let V be a r-dimensional subspace and $\bar{z}$ a point of R_k such that $E = V + \bar{z}$. Consider the measures λ_n and λ defined by

$$\lambda_n(A) = \mu_n(A + \bar{z}); \quad \lambda(A) = \mu(A + \bar{z}).$$

Since $\mu_n \Rightarrow \mu$ it follows that $\lambda_n \Rightarrow \lambda$. Also $\lambda_n(V) = \mu_n(E)$ and $\lambda(V) = \mu(E)$. Thus (b) implies that $\lambda_n(V) \to \lambda(V)$. Lemma 2.3.1. then implies that the measures $(\lambda_n)_V$ and $(\lambda)_V$ restricted to V and considered as measures on the vector space V are such that

$$(3.3.3) \qquad (\lambda_n)_V \Rightarrow \lambda_V.$$

Since μ is r-dimensional it follows that $\mu(H + \bar{z}) = \lambda(H) = 0$ for every hyperplane of the vector space V. Thus Theorem 3.3.1. applies to the sequence $(\lambda_n)_V$ and λ_V, and yields

$$\lim_{n \to \infty} \sup_{A \in \mathcal{B}_m} | \lambda_n(A \cap V) - \lambda(A \cap V)| = 0$$

for each m, where $\mathcal{B}_m$ is the class of all sets $A \subset R_k$ such that $A \cap V$ is the intersection of at most m half-spaces of the vector space V. It is easily verified that $\mathcal{H}_m \subset \mathcal{B}_{m+r}$. Consequently

$$\lim_{n \to \infty} \sup_{A \in \mathcal{H}_m} | \lambda_n(A \cap V) - \lambda(A \cap V)| = 0 \tag{3.3.4}$$

Since $\lambda(X - V) = \mu(X - E) = 0$, we have $\lambda_n(X - V) \to 0$. In view of this and the fact that $\mathcal{H}_m$ is invariant under translations, it now follows from (3.3.4) that

$$\lim_{n \to \infty} \sup_{A \in \mathcal{H}_m} | \mu_n(A) - \mu(A)| = 0.$$

This completes the proof.

Let μ be any arbitrary measure on R_k and let

$$\mu = \sum_{r=0}^{k-1} \sum_{i=1}^{\infty} \mu_r^{(i)} + \mu_k^{(0)} \tag{3.3.4}$$

be the decomposition asserted by the Lemma 3.2.3. Since all the measures $\mu_r^{(i)}$ and $\mu_k^{(0)}$ are all mutually orthogonal, there exists mutually disjoint sets $A_r^{(i)}$ and $A_k^{(o)}$ such that

$$\begin{aligned} \mu_r^{(i)}(B) &= \mu(B \cap A_r^{(i)}) \quad \text{and} \\ \mu_k^{(o)}(B) &= \mu(B \cap A_k^{(o)}) \end{aligned} \tag{3.3.5}$$

for each Borel set B.

Theorem 3.3.2. Let μ_n be a sequence of measures weakly converging to μ such that, in the notation of the above paragraph

(i) $\mu_n(A_r^{(i)}) \Rightarrow \mu(A_r^{(i)})$ for all r and i

(ii) $\mu_{nr}^{(i)} \Rightarrow \mu_r^{(i)}$ for all r and i, where $\mu_{nr}^{(i)}$ is defined by the relation $\mu_{nr}^{(i)}(B) = \mu_n(B \cap A_r^{(i)})$ for each Borel set B.

Then, for each m

$$\lim_{n \to \infty} \sup_{A \in \mathcal{F}_m} | \mu_n(A) - \mu(A) | = 0$$

Proof. Since $\mu_r^{(i)}$ is r-dimensional there exists a set $E_r^{(i)}$ which is a translate of a r-dimensional subspace such that $\mu_r^{(i)}(R_k - E_r^{(i)}) = 0$. We may suppose that $A_r^{(i)} \subset E_r^{(i)}$. Thus the condition (i) and (ii) of the above theorem imply that $\mu_{nr}^{(i)} \Rightarrow \mu_r^{(i)}$ and $\mu_{nr}^{(i)}(E_r^{(i)}) \Rightarrow \mu_r^{(i)}(E_r^{(i)})$. Therefore, by Lemma 3.3.1, we have

$$\lim_{n \to \infty} \sup_{A \in \mathcal{F}_m} | \mu_{nr}^{(i)}(A) - \mu_r^{(i)}(A) | = 0 \tag{3.3.6}$$

for $0 \leq r \leq k-1$ and all i. From Theorem 3.3.1 it follows that

$$\lim_{n \to \infty} \sup_{A \in \mathcal{F}_m} | \mu_{nk}^{(o)}(A) - \mu_k^{(o)}(A) | = 0 \tag{3.3.7}$$

Now define $E_N = \bigcup_{r=o}^{k-1} \bigcup_{i=1}^{N} A_r^{(i)} \cup A_k^{(o)}$. It then follows from assumption (i) of the theorem that $\mu_n(E_N) \Rightarrow \mu(E_N)$ as $n \Rightarrow \infty$ for each N. Or, what is the same thing,

$$\lim_{n \to \infty} \mu_n(R_k - E_N) = \mu(R_k - E_N) \tag{3.3.8}$$

It is easily seen that, for any set A,

$$(3.3.9) \qquad |\mu_n(A) - \mu(A)| \leq \sum_{r=0}^{k-1} \sum_{i=1}^{N} |\mu_{nr}^{(i)}(A) - \mu_r^{(i)}(A)| + |\mu_{nk}^{(0)}(A) - \mu_k^{(0)}(A)| + \mu_n(R_k - E_N) + \mu(R_k - E_N).$$

Take the supremum in (3.3.9) over the class $\mathcal{H}_m$ and then let $n \to \infty$. It then follows from (3.3.7) and (3.3.8) that

$$(3.3.10) \qquad \overline{\lim_{n \to \infty}} \sup_{A \in \mathcal{H}_m} |\mu_n(A) - \mu(A)| \leq 2\mu(R_k - E_N)$$

for each N. The assertion of the theorem follows from (3.3.10) by taking the limit as $N \to \infty$. This completes the proof.

4. Uniform convergence over the class $\mathcal{C}$ of all convex sets.

In the previous section it was shown (Theorem 3.3.1) that a mild continuity assumption on the limiting measure μ ensures uniformity of convergence over the class $\mathcal{H}_m$ for each m. It is natural to ask whether the same uniformity can be extended to m also i.e. to the class $\mathcal{H}_\infty = \bigcup_{m=1}^{\infty} \mathcal{H}_m$? If we look at the class $\mathcal{H}_\infty$ a little closely, we find that it is actually 'dense' in the class $\mathcal{C}$ of all convex sets, i.e. every set $C \in \mathcal{C}$ is the monotone limit of sets $\in \mathcal{H}_\infty$. Consequently, if $|\mu_n(A) - \mu(A)| \to 0$ uniformly for $A \in \mathcal{H}_\infty$, the same can be asserted for the class $\mathcal{C}$. Further, the class $\mathcal{C}$ is closed under monotone limits. Thus one is naturally lead to consider the class of all convex sets with regard to the problem of uniform convergence.

We begin with some remarks on convex sets. In the first place, a convex set need not be measurable*. However in the sequel by a convex set we shall always mean a measurable convex set. The symbol $\mathcal{C}$ will be used to denote the collection of all convex sets. In order not to interrupt the continuity of exposition here we have collected in a separate appendix all material on convex sets which will be used in this thesis, and reference may be made to this appendix for the terminology of convex sets and for their properties.

Theorem 3.4.1. Let μ be any measure on R_k for which $\mathcal{C}$ is a continuity class. If μ_n is a sequence of measures on R_k then $\mu_n \Rightarrow \mu$ if and only if

$$\lim_{n \to \infty} \sup_{A \in \mathcal{C}} | \mu_n(A) - \mu(A) | = 0 \qquad (3.4.1)$$

Proof. Let $S(o, m)$ be the sphere with centre at the origin and radius m. Then obviously $S(o, m)$ is a continuity set of the measure μ so that, for all $m > o$

$$\lim_{n \to \infty} \mu_n[S(o, m)] = \mu[S(o, m)].$$

Since the μ-measure of the set $R_k - S(o, m)$ can be made arbitrarily small by choosing m large it is clear that (3.4.1) will be proved if we can show that for each $m > o$

$$\lim_{n \to \infty} \sup_{A \in \mathcal{C}_m} | \mu_n(A) - \mu(A) | = 0 \qquad (3.4.2)$$

---

* Indeed let C be the interior and B the boundary of the unit sphere in R_k. If A is any non-measurable subset of B, then $C \cup A$ is convex but non-measurable.

where

$$\mathcal{C}_m = [C : C \in \mathcal{C} ; C \subset S(o, m)]$$

For an arbitrary convex set C, let C^o = interior of C and $\bar{C}$ = closure of C. Since $C^o \subset C \subset \bar{C}$ and $\mu(\bar{C} - C^o) = 0$, it is easy to deduce that

$$(3.4.3) \qquad |\mu_n(C) - \mu(C)| \leq \text{Max}\,(|\mu_n(\bar{C}) - \mu(\bar{C})|, |\mu_n(C^o) - \mu(C^o)|)$$

The above inequality shows that we need consider only convex sets which are either closed or open.

Now consider the class $\mathcal{C}_m^{(\delta)}$ defined by $\mathcal{C}_m^{(\delta)} = [C : C \in \mathcal{C}_m, r(C) \geq \delta]$ where $r(C)$ is the inradius of the convex set C. By combining (3.4.3), Theorem A.3.4 of the appendix and Theorem 2.5.1, it follows that for each $\delta > 0$ and all $m > o$,

$$(3.4.4) \qquad \lim_{n \to \infty} \sup_{A \in \mathcal{C}_m^{(\delta)}} |\mu_n(A) - \mu(A)| = 0$$

Next suppose that (3.4.2) is not true. This implies that there exists a sequence $C_1, C_2, \ldots$ of convex sets $\in \mathcal{C}_m$ and an $\varepsilon > 0$ such that

$$(3.4.5) \qquad |\mu_n(C_n) - \mu(C_n)| \geq \varepsilon > 0$$

It then clearly follows from (3.4.4) that $r(C_n) \to 0$ as $n \to \infty$. We may suppose, in virtue of Blaschke's theorem (cf. the Appendix), restricting ourselves to a subsequence if necessary, that there exists a closed set $C \in \mathcal{C}_m$ such that

$$(3.4.6) \qquad \lim_{n \to \infty} d(\bar{C}_n, C) = 0$$

where the distance function $d(\cdot, \cdot)$ is defined by (A.4.1) (cf. the Appendix). From the definition of the distance d(A, B) it follows that $C \subset S(\bar{C}_n, 2\varepsilon_n)$ where $\varepsilon_n = d(\bar{C}_n, C)$. Consequently $r(C) \leq r(\bar{C}_n) + 2\varepsilon_n = r(C_n) + 2\varepsilon_n$. Since both $r(C_n)$ and $\varepsilon_n \to 0$ it follows that $r(C) = 0$. i.e. C has empty interior. Thus there exist a hyperplane $H = [x: \ell(x) = p]$ such that $C \subset H$. Since $\mu(H) = 0$ choose $\eta > 0$ so small, that

$$\mu(A) < \frac{\varepsilon}{4}, \quad \text{where} \tag{3.4.7}$$

$$A = [x: \ p - \eta \leq \ell(x) \leq p + \eta]$$

From the definition of the distance $d(\cdot, \cdot)$ it is easily deduced that for large n

$$\bar{C}_n \subset S(C, \eta) \subset A .$$

Consequently,

$$\limsup_{n \to \infty} \mu_n(\bar{C}_n) \leq \lim_{n \to \infty} \mu_n(A) = \mu(A) \leq \frac{\varepsilon}{4} . \tag{3.4.8}$$

We also have

$$\limsup_{n \to \infty} \mu(C_n) \leq \mu(A) \leq \frac{\varepsilon}{4} . \tag{3.4.9}$$

Combining (3.4.8) and (3.4.9) we obtain

$$\limsup_{n \to \infty} | \mu_n(C_n) - \mu(C_n) | \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2} < \varepsilon .$$

This contradicts our assumption (3.4.5). The proof is thus complete.

Corollary to Theorem 3.4.1. If μ is a measure which is absolutely continuous with respect to the Lebesgue measure then $\mu_n \Rightarrow \mu$ if and only if

$$\lim_{n \to \infty} \sup_{A \in \mathcal{C}} | \mu_n(A) - \mu(A) | = 0$$

This follows from theorem 3.4.1 and the observation that the Lebesgue measure of the boundary of any convex set is zero.

We conclude this section with an example showing that the hypothesis of $\mathcal{C}$ being a continuity class cannot in general be omitted from Theorem 3.4.1. This example also shows that Theorem 3.2.1 cannot be improved to assert uniformity with respect to m also.

Let μ_n ($n = 0, 1, 2, \ldots$) be the Lebesgue measure of arcs concentrated on the circle

$$S_n = [(x, y) ; \quad x^2 + y^2 = r_n^2]$$

Let $r_n \searrow r_o$. Then it is easily verified that $\mu_n \Rightarrow \mu$ and the μ-measure of every hyperplane is zero. Therefore for each m

$$\lim_{n \to \infty} \sup_{A \in \mathcal{H}_m} |\mu_n(A) - \mu(A)| = 0$$

But there is no uniformity with regard to m. For consider the closed convex set

$$C = [(x, y); \quad x^2 + y^2 \leq r_o^2]$$

Then $\mu_n(C) = 0$ for all n and $\mu(C) = 2 \pi r_o$ so that

$$\sup_{A \in \mathcal{C}} | \mu_n(A) - \mu(A) | = 2 \pi r_o > 0$$

for all n.

3.5. Application to the convergence of continuous stochastic processes.

It is well known that a stochastic process can be thought of as a probability distribution in a function space. In many cases this function space can be made into a separable and complete metric space. Thus for instance we may consider distributions in $C[0, 1]$ - the space of all continuous functions on the interval $[0, 1]$, with the distance

$$\| x - y \| = \sup_{0 \le t \le 1} | x(t) - y(t) |$$

Some of the other function spaces that have been studied are $L_2[0, 1]$ - the space of all square integrable functions on $[0, 1]$ and $D[0, 1]$ - the space consisting of all functions on $[0, 1]$ with only discontinuities of the first kind. It is clear that results of Chapter 2 apply to all these spaces. This section contains a simple illustration in the case of the space $C[0, 1]$.

Let $\alpha(t)$ be a continuous function on $[0, 1]$. Consider the following two functionals on $C[0, 1]$ defined as follows:

$$\xi_\alpha(x) = \sup_{0 \le t \le 1} [x(t) - \alpha(t)]$$

(3.5.1)

$$\eta_\alpha(x) = \inf_{0 \le t \le 1} [x(t) - \alpha(t)]$$

for each $x = x(t) \in C[0, 1]$. It is clear that $\xi_\alpha(x)$ and $\eta_\alpha(x)$ are continuous functions on $C[0, 1]$. A subset $E \subset C[0, 1]$ of the form

$$E = [x: \alpha(t) \le x(t) \le \beta(t) \text{ for all } 0 \le t \le 1]$$

(3.5.2)

$$= [x: \xi_\beta(x) \le 0 ; \eta_\alpha(x) \ge 0]$$

is called a band with boundary functions $\alpha(t)$ and $\beta(t)$. Let $\mathcal{R}_m$ be the class of all continuous functions $\alpha(t)$ which are of the form

$$\alpha(t) = \int_0^t \gamma(t)\, dt \qquad \text{where} \tag{3.5.3}$$

$$\int_0^1 |\gamma(t)|^2\, dt \leq m$$

Let $\mathcal{B}_m$ be the class of all bands with both boundary functions belonging to the class $\mathcal{R}_m$.

Let W denote the distribution in C[0, 1] corresponding to the Brownian motion process. Actually the Brownian motion depends on a parameter σ^2 (Levy [26] chapter 6) where $\sigma^2 = E\, x^2(1)$. The distribution W may be supposed to correspond to the process with $\sigma = 1$. We begin with a lemma describing a class of continuity sets of W.

Lemma 3.5.1. For each function $\alpha(t)$ satisfying the conditions (3.5.3) the distributions $W\xi_\alpha^{-1}$ and $W\eta_\alpha^{-1}$ are both continuous (non-atomic). In particular $\mathcal{B}_m$ is a continuity class for W.

Proof. If x(t) denotes the Brownian motion process let W_α denote the distribution in C[0, 1] which corresponds to the process $y(t) = x(t) - \alpha(t)$. In other words the distribution W_α is a translation of W by α. i. e. $W_\alpha(A) = W(A+\alpha)$.

Now it is clear that for every real number a

$$W[\xi_\alpha(x) = a] = W_\alpha[\xi_0(x) = a] \tag{3.5.4}$$

where $\xi_0(x) = \sup_{0 \leq t \leq 1} x(t)$. We now utilize a result due to [illegible] and Skorohod ([33], p.426) generalizing an earlier result due to Cameran and Martin ([9]). This result asserts that when $\alpha(t)$

satisfies the condition (3.5.3) the distributions W_α and W are equivalent to each other i.e. $W(A) = 0$ when and only when $W_\alpha(A) = 0$. From this and (3.5.4) one obtains that $W[\xi_\alpha(x) = a] = 0$ if and only if

$$W[\xi_0(x) = a] = 0$$

A well known result in the theory of Brownian motion states that

$$W[\xi_0(x) \leq a] = \sqrt{\frac{2}{\pi}} \int_0^a e^{-t^2/2}\, dt$$

(see Doob [13], p. 392). Therefore $W[\xi_0(x) = a] = 0$ for all a. Thus Thus $W\xi_\alpha^{-1}$ has been shown to be a continuous distribution. By a similar argument $W\eta_\alpha^{-1}$ is also a continuous distribution. The proof of the lemma is thus complete.

Theorem 3.5.1. Let P_n, P $(n = 1, 2, \ldots)$ be distributions in $C[0, 1]$. Suppose that $P << W$ i.e. P is absolutely continuous with respect to W. Then $P_n \Rightarrow P$ and only if for each m

$$\lim_{n \to \infty} \sup_{A \in \mathcal{B}_m} |P_n(A) - P(A)| = 0 \tag{3.5.5}$$

Proof. Consider the class $\mathcal{F}_m$ of continuous maps of $C[0, 1]$ into the two-dimensional Euclidean space, of the form

$$T(x) = (\xi_\beta(x), \eta_\alpha(x)) \tag{3.5.6}$$

where $\alpha(t) \leq \beta(t)$ for $0 \leq t \leq 1$ and both the functions belong to $\mathcal{A}_m$. It is clear that the marginal distributions of PT^{-1} for each $T \in \mathcal{F}_m$ are continuous. i.e. the distributions $P\xi_\beta^{-1}$, $P\xi_\alpha^{-1}$ are both

continuous. This follows from the fact that by Lemma 3.4.1. both the distributions $W\xi_\beta^{-1}$ and $W\eta_\alpha^{-1}$ are continuous and the assumption $P << W$ implies that $P\xi_\beta^{-1} << W\xi_\beta^{-1}$ and $P\eta_\alpha^{-1} << W\eta_\alpha^{-1}$. Secondly the class $\mathcal{B}_m$ is contained in the class $= [A: A \subset C[0, 1]; A = T^{-1}(E)$ where $T \in \mathcal{F}_m$ and E a rectangle in $R_2]$. Thus the assertion (3.5.5) would be an immediate consequence of Theorem 2.5.1 if it could be shown that the family $\mathcal{F}_m$ is compact in the u.c.c. topology.

Consider the function $\alpha(t) \in \mathcal{A}_m$. By Schwartz inequality, one can easily deduce that

$$\text{(i)} \quad \| \alpha(t) \| \leq \sqrt{m}$$

$$\text{(ii)} \quad | \alpha(t_1) - \alpha(t_2) | \leq \sqrt{m} \, | t_1 - t_2 |^{\frac{1}{2}}$$

It then follows from Lemma (2.2.2) that the closure of $\mathcal{A}_m$ in $C[0, 1]$ is compact. Actually, $\mathcal{A}_m$ is a closed subset of $C[0, 1]$. To see this suppose $\alpha_n(t) \in \mathcal{A}_m$ $(n = 1, 2, \ldots)$ and $\alpha(t) \in C[0, 1]$ is such that

$$\text{(3.5.7)} \qquad \lim_{n \to \infty} \| \alpha_n(t) - \alpha(t) \| = 0.$$

By the definition of the class $\mathcal{A}_m$, there exists a sequence $\gamma_n(t)$ of functions such that

$$\text{(3.5.8)} \qquad \text{(i)} \quad \alpha_n(t) = \int_0^t \gamma_n(t)dt \text{ and}$$

$$\text{(3.5.9)} \qquad \text{(ii)} \quad \int_0^1 | \gamma_n(t) |^2 \, dt \leq m.$$

It follows from (3.5.9) that the sequence $\gamma_n(t)$, considered as elements of $L_2(0, 1)$, is norm bounded and hence is weakly conditionally compact.

i.e. there exists an element say $\gamma(t) \in L_2(0, 1)$ and a subsequence γ_{n_k} such that

$$\lim_{k \to \infty} \int_0^1 \gamma_{n_k}(t)\, g(t)dt = \int_0^1 \gamma(t)\, g(t)\, dt$$

for every $g(t) \in L_2(0, 1)$. In particular, for each t

$$\lim_{k \to \infty} \int_0^t \gamma_{n_k}(t)\, dt = \int_0^t \gamma(t)dt \tag{3.5.10}$$

It is clear that

$$\int_0^1 | \gamma(t) |^2 dt \leq m \tag{3.5.11}$$

Combining (3.5.7) - (3.5.11) it follows that $\alpha(t) \in \mathcal{R}_m$, i.e. the family $\mathcal{R}_m$ is a closed compact subset of $C[0, 1]$. Since the functionals $\xi_\alpha(x)$ and $\eta_\alpha(x)$ are also continuous when considered as functions of $\alpha \in C[0, 1]$ it is clear that compactness of $\mathcal{R}_m$ as a subset of $C[0, 1]$ implies that the family $\mathcal{F}_m$ is compact in the u.c.c. topology. The proof of the theorem is thus complete.

Remarks. 1. Although the above theorem considers only weak convergence of distributions to the Brownian motion process the method is also applicable to other stochastic processes realised in $C[0, 1]$. We only point out here that the special role of the Brownian process consists in the continuity of the distributions $W\xi_\alpha^{-1}$ and $W\eta_\alpha^{-1}$. In general, however, it is very difficult to ascertain the continuity or otherwise of $P\xi_\alpha^{-1}$ and $P\eta_\alpha^{-1}$. It may also be remarked that in some cases sufficient

conditions on α could be found such as those for instance where it is known that P_α - the translation of P by α is absolutely continuous with respect to P. The latter problem could be solved in the case of a large number of Gaussian Stochastic processes realised in $C[0, 1]$.

2. We had remarked earlier that results of the type obtained in this chapter might be useful in problems of estimation of the rapidity of convergence. An illustration is provided by an estimate due to Prohorov in the case of the approach to the Brownian motion process. Prohorov ([29] chapter 4) obtains an estimate of the difference $P_n - W$ where P_n is the distribution in $C[0, 1]$ corresponding to a random function $\xi(t)$ constructed from a series of independent observations $\xi_{n1}, \ldots \xi_{nn}$. The estimate referred to concerns the difference $P_n(H) - W(H)$ where H is a band with boundary functions satisfying the Lipschitz condition

$$| \alpha(t_1) - \alpha(t_2) | \leq K | t_1 - t_2 |$$

It is clear that every such H is in $\mathcal{B}_m$ and in fact $\mathcal{B}_m$ is a much bigger class. In view of theorem 3.5.1, it should be possible to find an estimate of $\sup_{A \in \mathcal{B}_m} | P_n(A) - W(A) |$. One might even conjecture that Prohorov's estimate is valid for this bigger class also. This however remains to be proved.

CHAPTER 4

THE LAW OF LARGE NUMBERS FOR GENERAL RANDOM FUNCTIONS, AND GENERALIZATIONS OF THE GLIVENKO-CANTELLI LEMMA

4.1. Introduction. This chapter is devoted to the study of laws of large numbers for general random functions. The last section of this Chapter concerns a generalization of the classical Glivenko-Cantelli lemma.

The Chapter consists of three parts. In the first part we investigate convergence properties of sample distribution functions. The main result (Theorem 4.3.2) is a generalization of an earlier result due to Fortet and Mourier. However the method used here depends on the results of Chapter 2 and is essentially different from theirs. This enables us to deduce immediately the laws of large numbers for random variables taking values in a separable Banach space. Our result is more precise than those of Mourier in that it identifies the limiting random element as a conditional expectation. We also define a conditional expectation for an L^1-valued random Banach space variable, and obtain sufficient conditions for its existence.

The second part of this Chapter is concerned with random variables taking values in D [0, 1] - the space of all functions on the interval [0, 1] with discontinuities (if any) only of the first kind. We obtain here laws of large numbers of both the 'almost sure 'and' mean of order α' types. Since D [0, 1] under the usual sup norm is a non-separable Banach space, we follow here a method which utilizes the topology for D [0, 1]

introduced by Skorohod. Though the results of this section could be obtained by generalizing the theorems on the convergence of sample distribution functions, we present a direct approach utilizing the structure of sets which are compact in the Skorohod topology.

The last part of the Chapter contains some results on the generalizations of the classical Glivenko-Cantelli lemma. The results of this section constitute a considerable improvement over the earlier results of Wolfowitz, Fortet and Mourier, and the recent result of Tucker.

4.2. Notation and definitions. Let (Ω, S, P) be a probability space. A real valued measurable function on Ω is called a random variable. ~~For a random variable.~~ For a random variable $f(w)$ on Ω we use the symbol $E f(w) = E f$ to denote the expected value of $f(w)$, $\int f(w)dP$. If $S_1 \subset S$ is a sub σ-field of S then $E(f/S_1)$ denotes the conditional expectation of the random variable $f(w)$ with respect to S_1.

Let X be an arbitrary separable and complete metric space, and $\mathcal{B}$ the class of Borel subsets of X. A mapping $\xi(w)$ of Ω into X is called an X-valued random variable (or a random element in X) if $\xi^{-1}(\mathcal{B}) \subset S$. The two special cases which will be considered in this Chapter are (i) $X =$ a separable Banach space, and (ii) $X = D[0, 1]$ the space of all functions on $[0, 1]$ with only discontinuities of the first kind. In both these cases the above definition of random elements in X is capable of other simpler interpretations. Some introductory remarks about random elements in $\mathcal{X}$ will be made here. The case (ii) of $D[0, 1]$-valued random variables will be considered in section 5.

We assume for purposes of this Chapter some of the standard results of Banach space theory. $\mathfrak{X}^*$ denotes the adjoint of $\mathfrak{X}$ i.e. the collection of all bounded linear functionals on $\mathfrak{X}$. Elements of $\mathfrak{X}^*$ will be denoted by x^*. The value of the linear functional x^* at the point x will be denoted by (x^*, x). To begin with it may be pointed out that if $\xi = \xi(w)$ is an $\mathfrak{X}$ -valued random variable, then $(x^*, \xi) = (x^*, \xi(w))$ is a random variable. Conversely, since $\mathfrak{X}$ is separable, if (x^*, ξ) is a random variable for each $x^* \in \mathfrak{X}^*$ then ξ itself is a random element in $\mathfrak{X}$ (see Hille and Phillips [20], Chapter III).

Let ξ be $\mathfrak{X}$ -valued random variable. Then the mean or expected value of ξ is defined to be the Pettis integral of $\xi(w)$ whenever it exists. More precisely, the expected value of ξ is an element $E\xi \in \mathfrak{X}$ which has the following property.

$$(x^*, E\xi) = E(x^*, \xi) \qquad \text{for each } x^* \in \mathfrak{X}^*$$

A particular situation of some importance where the existance of the expected value $E\xi$ is assured is the case when $E \| \xi \| < \infty$. (Here $\| x \|$ for $x \in \mathfrak{X}$ denotes the norm in $\mathfrak{X}$). For the above results we refer to Chapter III of the treatise of Hille and Phillips [20].

With regard to sequences of observations the most general such sequence that will be considered here is that of a strictly stationary sequence (briefly, an s.s. sequence). We may recall that a sequence $\{ \xi_n ; \quad n = 1, 2, \ldots \}$ is said to be an s.s. sequence of random elements in X provided for each pair of positive integers h and n, the distributions of $(\xi_1, \xi_2, \ldots \xi_n)$ and $(\xi_{n+1}, \ldots \xi_{n+h})$ are identical.

In particular, an independent and identically distributed sequence $\{\xi_n\}$ is strictly stationary.

A very convenient way of looking at a strictly stationary sequence is through a measure preserving transformation. Let T be a measure preserving transformation of Ω into itself and let $\xi = \xi(w)$ be a X-valued random variable. Then the sequence $\xi_n = \xi(T^n w)$ $(n = 0, 1, 2, \ldots)$ constitutes an s.s. sequence. As is well known we may, and in fact we do, assume that every s.s. sequence arises in this way. By the invariant σ-field associated with an s.s. sequence we mean the family $\mathcal{J}$ of all sets A such that $A \in S$ and $A = TA$.

The sample distribution function based on a sample $(\xi_1, \ldots \xi_n)$ of size n is defined to be the measure on X with masses $\frac{1}{n}$ at the points $\xi_1, \xi_2, \ldots \xi_n$. More precisely it is a function $\mu_n(A; w)$ defined for each Borel subset A of X and each sample point $w \in \Omega$ such that

$$\mu_n(A; w) = \frac{\chi_A(\xi_1) + \ldots + \chi_A(\xi_n)}{n}$$

where $\chi_A(x)$ is the characteristic function of the set A. It may be observed that the sample distribution function (s.d.f. in short) $\mu_n(A; w)$ has (trivially) the following two properties.

(i) for each fixed w, it is a measure in A

(ii) for each fixed A it is a measurable function in w.

As a matter of fact we will be considering in this Chapter more general functions $\mu(A; w)$ which possess the properties (i) and (ii) above. Such a function will be called a random measure.

In connection with the notion of a random measure introduced above the following observations are pertinent. If $f(x)$ is any measurable function on X then

$$\int_X f(x)\,\mu(dx\,;\,w) \tag{4.2.1}$$

is a random variable. In the sequel when there is no apparent risk of confusion we will abbreviate the right side of (4.2.1) to $\int f d\mu$.

We conclude this section with two lemmas on the existence of conditional distributions and conditional expectations.

Lemma 4.2.1. Let $S_1 \subset S$ be an arbitrary sub σ-field of S and ξ a X-valued random variable. Then there exists a random measure $\mu(A\,;\,w)$ such that, for every measurable function $f(x)$ on X for which $E\,|f(\xi)| < \infty$, we have

$$\int f(x)\,\mu(dx\,;\,w) = E[\,f(\xi)\,|\,S_1] \tag{4.2.2}$$

almost surely.

Proof. Since X is a separable and complete metric space it can be topologically imbedded as a G_δ in the countable product of unit intervals . (see Kelley,[21],p.207; Kuratowski,[24],p.337). Let Y be the countable product of unit intervals and let $\tilde{X}$ be the isomorphic image of X. Then obviously $\tilde{X}$ is a Borel subset of Y . In view of this isomorphism we may regard ξ as a random variable taking values in Y such that $\xi^{-1}(\mathcal{B}_Y) \subset S$, where $\mathcal{B}_Y$ stands for the Borel σ-field of Y. As is well known in this case a regular conditional distribution always

exists with respect to an arbitrary sub σ-field of S. (see Doob, [13]; p. 27-31 ; Loève, [27]; p. 361). This regular conditional distribution is precisely the random measure required by the lemma. The lemma is thus proved.

The next lemma is concerned with the existence of conditional expectation for random variables taking values in a separable Banach space $\mathfrak{X}$ Let ξ be such a random variable and let S_1 be an arbitrary sub σ-field of S. Then by Lemma 4.2.1 there exists a random measure $\mu(A\,;\ w)$ on $\mathfrak{X}$ which satisfies the equality (4.2.2).

Lemma 4.2.2. Suppose that $E\,\|\xi\| < \infty$. Then there exists an $\mathfrak{X}$-valued random element, which will be denoted by $E(\xi \mid S_1)$, such that for each $x^* \in \mathfrak{X}^*$, we have

$$(4.2.3) \qquad (x^*,\ E(\xi / S_1)) = E[(x^*\ ,\ \xi)/\ S_1],$$

almost surely. Moreover, we may suppose that with probability one, the following holds

$$(4.2.4) \qquad (x^*,\ \ E(\xi \mid S_1)) = \int (x^*,\ x)\ \mu(dx;\ w),$$

simultaneously for all $x^* \in \mathfrak{X}^*$.

Proof. Since $E\,\|\xi\| < \infty$, it follows that

$$(4.2.5) \qquad E(\|\xi\|/S_1) = \int \|x\|\ \mu(dx;\ w) \ < \infty$$

almost surely. Let Ω_o be the set of all w for which (4.2.5) is valid. Then $P(\Omega_o) = 1$. Thus we may assume that always $w \in \Omega_o$.

Consider the function $\Lambda(x^*\,;\,w)$ defined as follows

$$\Lambda(x^*\,;\,w) = \int (x^*,\,x)\,\mu(dx;\,w) \tag{4.2.6}$$

Then

$$|\,\Lambda(x^*\,;\,w)\,| \le \|\,x^*\,\| \int \|\,x\,\|\mu(dx;\,w),$$

and hence $\Lambda(x^*\,;\,w)$ is a bounded linear functional on $\mathfrak{X}^*$ for each fixed w. Also, for each fixed x^*, it follows from (4.2.6) and (4.2.2) that almost surely

$$\Lambda(x^*\,;\,w) = E[(x^*,\,\xi\,)\,|\,S_1]. \tag{4.2.7}$$

Next, we shall show that the functional $\Lambda(x^*;\,w)$ is actually continuous in the weak topology of $\mathfrak{X}^*$. Since $\mathfrak{X}$ is separable, the weak topology in $\mathfrak{X}^*$ induced by $\mathfrak{X}$ is a separable metric space. Suppose $x^*_n,\ x^* \in \mathfrak{X}^*$ are such that $x^*_n \to x^*$ weakly, i.e. for each $x \in \mathfrak{X}$

$$(x^*_n\,,\ x) \to (x^*\,,\,x) \tag{4.2.8}$$

Then, as is well known, this implies that there exists a constant $M < \infty$ such that

$$\|\,x^*_n\,\| \le M \quad \text{for all } n.$$

Thus, for all $x \in \mathfrak{X}$, we have

$$|(x^*_n\,,\,x)| \le M\ \|\,x\,\|. \tag{4.2.9}$$

It follows from (4.2.8), (4.2.9) and the bounded convergence theorem that

for each $w \in \Omega_o$,

$$\lim_{n \to \infty} \Lambda(x_n^* ; w) = \Lambda(x^* ; w)$$

i.e. for each $w \in \Omega_o$, $\Lambda(x^* ; w)$ is continuous in the weak topology of $\mathcal{X}^*$.

According to a well known theorem, a bounded linear functional on $\mathcal{X}^*$ which is continuous in the weak topology corresponds to an element of $\mathcal{X}$. It follows hence that there exists an element of $\mathcal{X}$ say $\xi_1(w)$ corresponding to each $w \in \Omega_o$ such that

$$(4.2.10) \qquad (x^*, \xi_1(w)) = \Lambda(x^* ; w) .$$

Since $\Lambda(x^* ; w)$ is a measurable function of w for each x^* it follows from (4.2.10) that $\xi_1(w)$ is actually a random element in $\mathcal{X}$. Denote this random element by $E(\xi / S_1)$. Thus for each $w \in \Omega_o$, one has simultaneously for all $x^* \in \mathcal{X}^*$

$$(4.2.11) \qquad (x^*, E(\xi / S_1)) = \Lambda(x^* ; w) = \int (x^*, x)\, \mu(dx ; w)$$

This proves (4.2.4). (4.2.7) together with (4.2.11) implies (4.2.3). This completes the proof of the lemma.

4.3. Convergence properties of sample distribution functions. In this section several results on the convergence of the s.d.f. are obtained. Here we utilize the results of Chapters 1 and 2 on the relation between weak convergence and uniform convergence. The theorem which connects these problems with the results and methods of Part I is a result due to

Varadarajan (theorem 4.3.1).

Throughout this section, $\{ \xi_n ; \ n = 1, 2, \ldots \}$ is an s.s. sequence, and $\mu_n (A ; w)$ is the associated sequence of sample distribution functions. Birkhoff's individual ergodic theorem (Loeve [27]; p. 421) states that for every function $f(x)$ on X for which $E | f(\xi_1)| < \infty$, one has almost surely

$$\lim_{n \to \infty} \frac{f(\xi_1) + \ldots + f(\xi_n)}{n} = E[f(\xi_1) | \mathcal{J}]$$

where $\mathcal{J}$ is the invariant σ-field. In terms of s.d.fs this may be rewritten as

$$(4.3.1) \qquad \lim_{n \to \infty} \int f(x) \mu_n (dx ; w) = E[f(\xi_1) | \mathcal{J}]$$

almost surely. Again, by Lemma 4.2.1 there exists a random measure $\mu(A;w)$ such that almost surely

$$(4.3.2) \qquad E[f(\xi_1) | \mathcal{J}] = \int f(x) \mu(dx ; w)$$

Throughout this section, the random measure $\mu(A; w)$ is assumed to be determined by (4.3.2).

With these preliminary remarks we can now state our first theorem which is essentially due to Varadarajan ([37]; p. 24). As observed earlier this theorem is the connecting link with the results and methods of Part I.

Theorem 4.3.1. With probability one, the following holds

(4.3.3) $$\mu_n \Rightarrow \mu$$

This theorem is an immediate consequence of (4.3.1), (4.3.2) and Theorem 2.3.2 of Chapter 2.

Remark. The following is a slight modification of the above theorem. Let $B \in \mathcal{B}$ be any fixed Borel set. Let $\nu_n(A ; w)$ and $\nu(A ; w)$ be the random measures defined by

$$\nu_n(A ; w) = \mu_n(A \cap B ; w) ; \quad \nu(A ; w) = \mu(A \cap B; w)$$

Then

(4.3.4) $$P[w ; \ \nu_n \Rightarrow \nu \ ] = 1.$$

The proof is similar to that of Theorem 4.3.1.

The following lemma will be useful in deducing L_p - convergence theorems from almost everywhere convergence.

Lemma 4.3.1. Let $\mathcal{R}$ be any class of measurable functions on X such that

(i) there exists a measurable function $g(x)$ on X such that $|f(x)| \leq g(x)$ for all $x \in X$ and all $f \in \mathcal{R}$.

(ii) $E[g(\xi_1)]^{1+\alpha} < \infty$ for some $\alpha \geq 0$

Suppose that $\lim_{n \to \infty} P[\eta_n > \varepsilon] = 0$, where η_n is defined by

(4.3.5) $$\eta_n = \sup_{f \in \mathcal{R}} \left| \int f d\mu_n - \int f d\mu \right|$$

Then

(4.3.6) $$\lim_{n \to \infty} E \eta_n^{1+\alpha} = 0$$

Proof. Let h_n, h be random variables defined by

$$h_n = \int [g(x)]^{1+\alpha} \, d\mu_n$$

$$h = \int [g(x)]^{1+\alpha} \, d\mu$$

Then

(a) $h_n \geq 0$, $h \geq 0$;

(b) $E\, h_n = E\, h$ for $n = 1, 2, 3, \ldots$;

(c) $P[\, h_n \to h \,] = 1$.

Properties (a) and (b) are trivial, while (c) is a consequence of assumption (ii) and Birkhoff's individual ergodic theorem. We conclude from the above three properties, by application of a theorem of Scheffe ([31], p. 435) that

$$\lim_{n \to \infty} E\, |\, h_n - h\, | = 0 \tag{4.3.7}$$

For any $f \in \mathcal{A}$, we have

$$|\int f d\mu_n - \int f d\mu\,|^{1+\alpha} \leq 2^{1+\alpha}[\,|\int f d\mu_n\,|^{1+\alpha} + |\int f d\mu\,|^{1+\alpha}\,]$$

$$\leq 2^{1+\alpha}\,[(\int g d\mu_n)^{1+\alpha} + (\int g d\mu)^{1+\alpha}]$$

$$\leq 2^{1+\alpha}\,(h_n + h).$$

The last step in (4.3.7) is by means of Holder's inequality. Thus

$$\eta_n^{1+\alpha} \leq 2^{1+\alpha}\,(h_n + h). \tag{4.3.8}$$

Now let E_n denote the set $= [\eta_n \geq \varepsilon]$. By assumption, $P(E_n) \to 0$ as $n \to \infty$. Combining (4.3.8), (4.3.7), and the fact that $\lim_{n \to \infty} \int_{E_n} h dP = 0$, it follows that

$$\lim_{n \to \infty} \int_{E_n} \eta_n^{1+\alpha} dP = 0 . \tag{4.3.9}$$

Since

$$E \eta_n^{1+\alpha} \leq \varepsilon^{1+\alpha} + \int_{E_n} \eta_n^{1+\alpha} dP ,$$

it follows from (4.3.9) that

$$\lim_{n \to \infty} E \eta_n^{1+\alpha} = 0 .$$

The proof is thus complete.

The main theorem of this section on the convergence of s.d.f's can now be stated as follows.

Theorem 4.3.2. Let $\mathcal{R}$ be a family of continuous functions on X satisfying the following conditions:

(i) There exists a continuous function $g(x)$ on X such that $|f(x)| \leq g(x)$ for all $x \in X$ and $f \in \mathcal{R}$.

(ii) $E g(\xi_1) < \infty$.

(iii) $\mathcal{R}$ is compact in the topology of uniform convergence on compacta.

Let $\eta_n = \sup_{f \in \mathcal{R}} |\int f d\mu_n - \int f d\mu|$. Then

$$P[\lim_{n \to \infty} \eta_n = 0] = 1. \tag{4.3.10}$$

If, in addition to the above assumptions, $E[g(\xi_1)]^{1+\alpha} < \infty$ for some

$\alpha \geq 0$; then

$$\lim_{n \to \infty} E\,\eta_n^{1+\alpha} = 0. \tag{4.3.11}$$

Proof. From Theorem 4.3.1 and Birkhoff's individual ergodic theorem applied to the sequence $g(\xi_n)$, one obtains, that with probability one, the following two assertions hold.

$$\text{(a)} \quad \mu_n \Rightarrow \mu$$
$$\text{(b)} \quad \int g\,d\mu_n \Rightarrow \int g\,d\mu. \tag{4.3.12}$$

(4.3.10) is an immediate consequence of (4.3.12) and Theorem 2.4.2, of Chapter 2. The second part of the theorem follows readily from Lemma (4.3.1) and the assertion (4.3.10).

This completes the proof.

Remarks. Results similar to the above theorem have been obtained by Fortet and Mourier [16]. They consider the class of functions $\mathcal{F}$ which satisfy a type of Lipschitz condition viz

$$| f(x_1) - f(x_2) | \leq M\, d(x_1, x_2)$$

where the function $d(x_1, x_2)$ is the metric in X. Such a family is conditionally compact in the u.c.c. topology as is easily seen from Lemma 2.2.1 of Chapter 2. Thus it is clear that the results of [16] could be obtained as a particular case from our results. The method employed in [16] is also different from ours. As a matter of fact our main purpose in obtaining Theorem 4.3.2 is to apply it to deduce laws of large numbers for general random elements. On the other hand Fortet and Mourier deduce their results

on the convergence of s.d.fs from the law of large numbers.

4.4. Laws of large numbers for $\mathfrak{X}$-valued random variables. In this section we utilize the results of Section 3 to deduce the laws of large numbers for random elements with values in a separable Banach space. Let $\xi_1, \xi_2, \ldots$ be an s.s. sequence of random elements in $\mathfrak{X}$ and let $\mu_n(A\,;\,w)$ denote their corresponding s.d.fs. Let

$$S_n = \frac{\xi_1 + \ldots + \xi_n}{n}$$

Let $\mu(A\,;\,w)$ be the random measure such that $E[\,f(\xi_1)\mid \mathcal{J}\,] = \int f d\mu$ almost surely for each f such that $E\mid f(\xi_1)\mid < \infty$. If we suppose that $E\,\|\xi_1\| < \infty$, then Lemma 4.2.2 assures the existence of random element $E(\xi_1 \mid \mathcal{J})$ such that, almost surely, one has simultaneously for all $x^* \in \mathfrak{X}^*$

$$(4.4.1) \qquad (x^*, E(\xi_1 \mid \mathcal{J})) = \int (x^*, x)\, d\mu.$$

Theorem 4.4.1. Suppose that $E\,\|\xi_1\| < \infty$. Then, with probability one,

$$\lim_{n \to \infty} \|S_n - E(\xi_1 \mid \mathcal{J})\| = 0.$$

If we assume in addition that $E\,\|\xi_1\|^{1+\alpha} < \infty$ for some $\alpha \geq 0$, then

$$\lim_{n \to \infty} E\,\|S_n - E(\xi_1 \mid \mathcal{J})\|^{1+\alpha} = 0.$$

Proof. It is clear that

$$(x^*, S_n) = \int (x^*, x)\, d\mu_n$$

for all $x^* \in \mathfrak{X}^*$. Consequently, from (4.4.1) it follows that, almost surely one has, simultaneously for all $x^* \in \mathfrak{X}^*$

more intimate use of the structure of the space D[0, 1] itself.

We begin with some preliminary material on the space D[0, 1]. The space D[0, 1] is the collection of all functions $x(t)$ on the interval [0, 1] which possess the following properties.

(i) for each t, $0 < t < 1$ the right and left limits $x(t + 0) = \lim_{s \downarrow t} x(s)$ and $x(t - 0) = \lim_{s \uparrow t} x(s)$ both exist,

(ii) $\lim_{s \downarrow 0} x(s)$ exists and $= x(0)$, and $\lim_{s \uparrow 1} x(s)$ exists and $= x(1)$,

(iii) for each $0 < t < 1$, $x(t + 0) = x(t)$.

Obviously D[0, 1] is a vector space. D[0, 1] can be normed into a Banach space with the norm

$$\| x \| = \sup_{0 \leq t \leq 1} | x(t) | ,$$

but this space is however nonseparable. The non separability of the space D[0, 1] is a great handicap, since it renders many of the methods of classical functional analysis (especially those regarding measures) inapplicable. One of the methods proposed to over come this difficulty is to consider other topologies for the space D[0, 1] which take into account the nature of functions entering the space D[0, 1]. This was accomplished by Skorohod ([32]), who introduced for this purpose four different topologies in D[0, 1]. Here we will be utilizing one of these topologies, which Skorohod calls the J_1 topology. The properties of D[0, 1] endowed with this topology which we will require are given in the following paragraphs. We will call this topology the Skorohod topology, or S-topology. For a complete description

of this topology see Skorohod [32], Kolmogorov [22], and also Prohorov ([29]; Appendix II).

The space $D[0, 1]$ endowed with the S-topology becomes a separable metric space which is topologically complete*. The structure of compact subsets of $D[0, 1]$ in the S-topology is known. We employ here the following characterization given by Kolmogorov ([22], p. 244):

Theorem 4.5.1. A subset $K \subset D[0, 1]$ is compact in the S-topology if and only if the following conditions are fulfilled.

(i) for each t, $0 \leq t \leq 1$

$$\sup_{x \in K} |x(t)| < \infty$$

(ii) for every $\varepsilon > 0$, there exist points $0 = t_0 < t_1 < t_2 \ldots < t_k = 1$ such that

$$\sup_{t_i \leq t', t'' < t_{i+1}} |x(t') - x(t'')| < \varepsilon$$

for $i = 0, 1, 2 \ldots k-1$ and all $x = x(t) \in K$.

The S-topology is not compatible with the vector structure of $D[0, 1]$, i.e. if the sequences $x_n(t), y_n(t) \in D[0, 1]$ for $(n = 0, 1, 2 \ldots)$ are such that

$$x_n(t) \xrightarrow{S} x_0(t)$$

$$y_n(t) \xrightarrow{S} y_0(t)$$

as $n \to \infty$, where the convergence is in the Skorohod topology, then it is not necessarily true that

---

* A metric space X is topologically complete if there is an equivalent metric which converts X into a complete metric space.

$$x_n(t) + y_n(t) \xrightarrow{S} x_0(t) + y_0(t)$$

This will not, however, affect us as we make only an indirect use of the S-topology; the topology itself never enters in our theorems.

The projections i.e. the functionals $f_t(x)$ on $D[0, 1]$ defined by

$$f_t(x) = x(t) \qquad \text{for } x \in D[0, 1]$$

are not continuous in the S-topology. However, they are all measurable with respect to the Borel σ-field $\mathcal{B}$ of the S-topology of $D[0, 1]$. The functional

$$\| x \| = \sup_{0 \leq t \leq 1} | x(t) |$$

is also Borel measurable. (In fact, the functional $\| x \|$ is continuous in the S-topology)

We now consider random elements in $D[0, 1]$. As in our earlier definition, here also by a $D[0, 1]$ - valued random variable we mean a mapping $\xi(w)$ (or $\xi(t, w)$) into $D[0, 1]$ such that $\xi^{-1}(\mathcal{B}) \subset S$, where $\mathcal{B}$ is the Borel σ-field of S-topology of $D[0, 1]$. The dependence of this definition on the topology we employ is only an apparent one, and the following definition is equivalent to the one just given : A mapping $\xi(t, w)$ into $D[0, 1]$ is a $D[0, 1]$-valued random variable if and only if for each t, $0 \leq t \leq 1$ the function $\xi(t, w)$ is a random variable. Without going into details we remark here that this equivalence is a consequence of the easily established fact that the Borel σ-field $\mathcal{B}$ is precisely the minimal σ-field with respect to which all the functionals $f_t(x)$ are measurable.

Let $\xi_1(t, w)$, $\xi_2(t, w)$, ... , $\xi_n(t, w)$ ... , be a strictly stationary sequence of random elements in $D[0, 1]$, and define

$$S_n(t, w) = \frac{\xi_1(t, w) + \ldots + \xi_n(t, w)}{n}$$

Suppose that

$$(4.5.1) \qquad E \| \xi_1 \| = E\left[\sup_{0 \le t \le 1} | \xi_1(t, w)| \right] < \infty .$$

For each t, $0 \le t \le 1$ consider the random variable

$$(4.5.2) \qquad m(t, w) = E[\xi_1(t, w)/\mathcal{J}]$$

where $\mathcal{J}$ is the invariant σ-field. On the strength of the assumption (4.5.1), it is not difficult to deduce (as in Lemma 4.2.2) that for almost all w, $m(t, w) \in D[0, 1]$ i.e. $m(t, w)$ is a random element in $D[0, 1]$.

Theorem 4.5.2. With probability one,

$$(4.5.3) \qquad \lim_{n \to \infty} \sup_{0 \le t \le 1} | S_n(t, w) - m(t, w) | = 0$$

If we suppose in addition that $E \| \xi_1 \|^{1+\alpha} < \infty$ for some $\alpha \ge 0$ then

$$(4.5.4) \qquad \lim_{n \to \infty} E\left[\sup_{0 \le t \le 1} | S_n(t, w) - m(t, w)|^{1+\alpha} \right] = 0$$

Proof. Let $\mu_n(A, w)$ be the sample distribution function of the sequence $\{\xi_n\}$, and let the random measure $\mu(A, w)$ be defined by the equality

$$E[f(\xi_1)/\mathcal{J}] = \int f(x)\, \mu(dx, w)$$

for each function f for which $E \mid f(\xi_1) \mid < \infty$. (cf. Lemma 4.2.2). In particular, one has for each t, $0 \leq t \leq 1$

$$m(t, w) = \int f_t(x)\, \mu(dx, w)$$

almost surely, where for each $x \in D[0, 1]$, $f_t(x) = x(t)$.

Define for each t, $0 \leq t \leq 1$

$$\varepsilon_n(t) = \mid S_n(t, w) - m(t, w) \mid$$

$$= \mid \int f_t(x)\, \mu_n(dx, w) - \int f_t(x)\, \mu(dx, w) \mid .$$

Then from Birkhoff's individual ergodic theorem it follows that for each t, $0 \leq t \leq 1$, the following holds:

$$(4.5.5) \qquad P\left[\lim_{n \to \infty} \varepsilon_n(t) = 0\right] = 1.$$

Let K be an arbitrary subset of $D[0, 1]$ which is compact in the S-topology. Let $\varepsilon > 0$ be arbitrary. Then, according to Theorem 4.5.1, there exist points

$$(4.5.6) \qquad 0 = t_0 < t_1 < \ldots < t_{k-1} < t_k = 1$$

(depending only on ε and K) such that, for each $x \in K$,

$$(4.5.7) \qquad \sup_{t_i \leq t', t'' < t_{i+1}} \mid x(t') - x(t'') \mid \leq \varepsilon$$

for $i = 0, 1, 2, \ldots k-1$.

Let t be an arbitrary point such that $t_i \leq t < t_{i+1}$ and consider the function $\varepsilon_n(t)$. It is easily verified that

$$\varepsilon_n(t) \leq \left| \int f_{t_i}(x)\, d\mu_n - \int f_{t_i}(x)\, d\mu \right| +$$

$$+ \int_K | f_t(x) - f_{t_i}(x) |\, d\mu_n + \int_K | f_t(x) - f_{t_i}(x) |\, d\mu +$$

$$+ 2 \int_{K'} \| x \|\, d\mu_n + 2 \int_{K'} \| x \|\, d\mu,$$

where K' denotes the complement of K in $D[0, 1]$. If we utilize (4.5.7) this simplifies to

$$(4.5.8) \qquad \sup_{t_i \leq t < t_{i+1}} \varepsilon_n(t) \leq \varepsilon_n(t_i) + 2\varepsilon + 2 \int_{K'} \| x \|\, d\mu_n +$$

$$+ 2 \int_{K'} \| x \|\, d\mu.$$

Since $E \| \xi_1 \| < \infty$, it follows from Birkhoff's individual ergodic theorem that

$$(4.5.9) \qquad \lim_{n \to \infty} \int_{K'} \| x \|\, d\mu_n = \int_{K'} \| x \|\, d\mu$$

almost surely. Combining (4.5.5), (4.5.8) and (4.5.9), it follows that with probability one the following inequality holds:

$$\overline{\lim_{n \to \infty}} \sup_{t_i \leq t < t_{i+1}} \varepsilon_n(t) \leq 2\varepsilon + 4 \int_{K'} \| x \|\, d\mu$$

for each $i = 0, 1, \ldots k - 1$. Or, what is the same thing,

$$(4.5.10) \qquad \overline{\lim_{n \to \infty}} \sup_{0 \le t \le 1} \varepsilon_n(t) \le 2\varepsilon + 4 \int_{K'} \| x \| \, d\mu$$

Thus the inequality 4.5.10 holds almost surely for each compact subset of $D[0, 1]$.

It is well known that every finite measure in a separable and complete metric space is concentrated in a σ-compact subset of $D[0, 1]$. It now follows that

$$(4.5.11) \qquad \inf_{K} \int_{K'} \| x \| \, d\mu = 0$$

almost surely, where the infimum in (4.5.11) is taken over all compact subsets of $D[0, 1]$. Since ε is arbitrary (4.5.10) and (4.5.11) together yield the assertion (4.5.3) of the theorem.

The second part of theorem is then an immediate consequence of (4.5.3) and lemma 4.3.1.

The proof of the theorem is thus complete.

4.6. Generalizations of the Glivenko - Cantelli lemma. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent and identically distributed real valued random variables. If $F_n(x)$ and $F(x)$ denote, respectively, the s.d.f. and true d.f. then

$$(4.6.1.) \qquad P\left[\sup_{-\infty < x < +\infty} | F_n(x) - F(x) | \to 0 \right] = 1 .$$

This is the Glivenko-Cantelli lemma. In the introductory Chapter 1 we indicated how this might be interpreted as a law of large numbers for certain random elements in $D[0, 1]$. The results of section 5 contain a complete

generalization of (4.6.1) in this direction.

Several authors have obtained generalization of (4.6.1) in a different direction. This consists in interpreting (4.6.1) as asserting uniform convergence over a certain class of sets. Taking the ξ's to be random vectors (say k-dimensional) Wolfowitz ([41]) has proved the following result: with probability one

$$(4.6.2) \qquad \lim_{n \to \infty} \sup_{H} | \mu_n(H) - \mu(H) | = 0$$

where the supremum is taken over the class of all half spaces and μ_n, μ stand for the s.d.f. and the common distribution respectively (see also Fortet and Mourier [16]). Blum ([8]) has shown that (4.6.2) could be improved to assert uniform convergence over a bigger class. Recently Tucker ([34]) extended (4.6.1) to the case when the sequence $\xi_1, \xi_2,$ is only known to be strictly stationary. In this section we apply the general theorems of Part I to obtain results which constitute an improvement of all these results.

Let $\xi_1, \xi_2, \ldots$ be a strictly stationary sequence of k-dimensional random vectors, and let μ_n denote the corresponding sequence of s.d.fs. Let the random measure $\mu = \mu(A, w)$ be defined by the relation

$$\mu(A, w) = P[\xi_1 \in A / \mathcal{I}]$$

(see Lemma 4.2.1). Let $\mathcal{H}_m$ denote the class of all sets which are intersections of at most m half-spaces.

Theorem 4.6.1. For each positive integer m,

$$(4.6.3) \qquad \lim_{n \to \infty} \sup_{A \in \mathcal{H}_n} | \mu_n(A) - \mu(A) | = 0$$

with probability one.

Proof: Let λ denote the distribution of ξ_1 i.e.

$$\lambda(A) = P [\xi_1 \in A]$$

By Lemma 3.2.3, there exist mutually orthogonal measures $\lambda_r^{(i)}$ such that

$$(4.6.4) \qquad \lambda = \sum_{r=0}^{k-1} \sum_{i=0}^{\infty} \lambda_r^{(i)} + \lambda_k^{(0)} .$$

Let $A_r^{(i)}$ ($r = 0, 1 \ldots k-1$, $i = 0, 1, 2, \ldots$; and $r = k$, $i = 0$) be mutually disjoint Borel sets such that the measures $\lambda_r^{(i)}$ is defined by

$$(4.6.5) \qquad \lambda_r^{(i)}(E) = \lambda (E \cap A_r^{(i)})$$

Consider the random measures $\mu(A, w)$, $\mu_r^{(i)}(A, w)$ defined by

$$(4.6.7) \qquad \begin{aligned} \mu(A, w) &= P [\xi_1 \in A \mid \mathcal{J}] \\ \mu_r^{(i)}(A, w) &= P [\xi_1 \in A \cap A_r^{(i)} \mid \mathcal{J}] \end{aligned}$$

The existence of such measures is guaranteed by Lemma 4.2.1. It follows easily from (4.6.4), (4.6.5) and (4.6.7) that almost surely

$$(4.6.8) \qquad \mu(A, w) = \sum_{r=0}^{k-1} \sum_{i=0}^{\infty} \mu_n^{(i)}(A, w) + \mu_k^{(0)}(A, w).$$

It is not difficult to see from (4.6.5) and (4.6.7) that the measure $\mu_r^{(i)}$ is the restriction of μ to the set $A_r^{(i)}$, i.e. with probability one,

$$\mu_r^{(i)}(E, w) = \mu(E \cap A_r^{(i)}, w) \tag{4.6.9}$$

simultaneously for all Borel sets E, and all r and i.

Now consider the sequence of s.d.fs $\mu_n(A, w)$. Let the measures $\mu_{nr}^{(i)}$ be defined as follows. For each Borel set E

$$\mu_{nr}^{(i)}(E) = \mu_n(A_r^{(i)} \cap E, w)$$

for all r and i. With this definition we can deduce from Theorem 4.3.1 (see also the remark following that theorem; in particular 4.3.4) that with probability one

$$\begin{aligned} &\text{(i)} \quad \mu_n \Rightarrow \mu \\ &\text{(ii)} \quad \mu_{nr}^{(i)} \Rightarrow \mu_r^{(i)} \quad \text{for } r = 0, 1 \ldots k-1 \text{ and all } i \\ &\text{(iii)} \quad \mu_{nk}^{(0)} \Rightarrow \mu_k^{(0)} \end{aligned} \tag{4.6.10}$$

From Birkhoff's individual ergodic theorem it follows that with probability one

$$\lim_{n \to \infty} \mu_n(A_r^{(i)}, w) = \mu(A_r^{(i)}, w) \tag{4.6.11}$$

for $r = 0, 1 \ldots k-1$ and all i, and $r = k$, $i = 0$. Thus (4.6.10), (4.6.11) and Theorem 3.3.2 together imply that, with probability one, we have

$$\lim_{n \to \infty} \sup_{A \in \mathfrak{H}_m} | \mu_n(A) - \mu(A) | = 0.$$

for each integer m. This completes the proof.

Our next result is a refinement of Theorem 4.6.1 in certain special cases. We replace the class $\mathcal{H}_m$ by the class $\mathcal{C}$ of all convex sets.

Theorem 4.6.2. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent and identically distributed random vectors, and let μ_n and μ denote their s.d.f and common actual d.f., respectively. Suppose the measure μ is such that $\mathcal{C}$ is a continuity class for μ. Then, with probability one,

$$\lim_{n \to \infty} \sup_{A \in \mathcal{C}} | \mu_n(A) - \mu(A) | = 0 . \tag{4.6.12}$$

Proof. This theorem is an immediate consequence of Theorems 4.3.1 and 3.4.1.

Remark. In general it is impossible to drop the continuity assumption made in Theorem 4.6.2. This is shown by the example given below. The example also shows that Theorem 4.6.1 cannot be improved to assert uniform convergence over the class $\bigcup_m \mathcal{H}_m$. In this connection we may draw attention to the remarks made in Section 4 of Chapter 3.

Example. Let S denote the surface of the unit sphere in R_k and let S_0 be the interior of the sphere. Let μ be any non atomic probability distribution concentrated on the surface S. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random vectors with μ as their common distribution. If μ_n denotes their s.d.fs then with probability one

$$\lim_{n \to \infty} \sup_{A \in \mathcal{H}_m} | \mu_n(A) - \mu(A) | = 0$$

for each integer m, by Theorem 4.6.1. Now suppose A is any Borel subset of S. It is easy to see that $S_0 \cup A$ is convex. Consequently,

$$\sup_{A \in \mathcal{C}} | \mu_n(A) - \mu(A) | = \sup_{A \in \mathcal{B}} | \mu_n(A) - \mu(A) |$$

where $\mathcal{B}$ is the class of all Borel subsets of S. Since μ_n and μ are concentrated in S and μ_n is atomic, while μ is not, it is clear that we always have

$$\sup_{A \in \mathcal{B}} | \mu_n(A) - \mu(A) | = 1.$$

Thus Theorem 4.6.2 is not valid in this case.

CHAPTER 5

On the Central Limit Theorem in R_k. Estimation of the remainder term.

5.1. Introduction. Let $Z^{(1)}, Z^{(2)}, \ldots$ be a sequence of independent and identically distributed random vectors in R_k with mean vector zero and variance covariance matrix Σ. If P_n is the probability distribution of the normed sum

$$Y^{(n)} = \frac{Z^{(1)} + \ldots + Z^{(n)}}{\sqrt{n}}$$

and $\Phi(0, \Sigma)$ that of the corresponding normal distribution, then the analogue of the classical central limit theorem states that

$$P_n \Rightarrow \Phi(0, \Sigma).$$

Questions that naturally arise here are how large is the error $| P_n - \Phi(0, \Sigma) |$, and what is the asymptotic expansion of the difference $P_n - P$. These questions have been thoroughly investigated in the case $k = 1$. In this Chapter we study these questions for the general case $k > 1$ and obtain solutions to both these problems.

A feature of this study is that the error is investigated for a large class of sets. The class considered is the one suggested by Theorem 3.4.1. of Chapter 3, i.e. the class of all convex subsets of R_k.

The accuracy of normal approximation in R_k has been studied previously by some authors. Esseen ([15], Chapter 7) studies the error for sets A of the form

$$A = [\, \underset{\sim}{x} : \ \underset{\sim}{x}' \Sigma^{-1} \underset{\sim}{x} \leq r \,]$$

for real $r > 0$ and shows that

$$(5.1.1) \qquad | P_n(A) - \Phi(A) | \leq c(k)\, \beta_4^{-3/2}\, n^{-k/(k+1)}$$

where $c(k)$ is a constant depending only on k and β_4 is defined by (5.2.2). (The result is stated here for the case $\Sigma = I$ - the identity matrix). Another important work is that of Bergström ([3]) who studies the difference of the distribution functions $F_n(x)$ and $\Phi(x)$ of P_n and P respectively. He shows that

$$(5.1.2) \qquad | F_n(x) - \Phi(x) | \leq \frac{c(k)}{\sqrt{n}}\, \beta_3 \max_{i,j} | \sigma^{ij} |^{3/2}$$

where $\{ \sigma^{ij} \}$ are the elements of Σ^{-1}. Bergström has also studied the problem for the non-identically distributed case (cf. [4]).

The problem of asymptotic expansion does not seem to have received any attention in the literature.

The first main result obtained in this Chapter is the following

$$(5.1.3) \qquad \sup_{A} | P_n(A) - \Phi(A) | \leq c(k)\, \beta_4^{3/2}\, \frac{(\log n)^{\alpha}}{\sqrt{n}}$$

where the supremum is taken over the class $\mathcal{C}$ all convex sets A,

$$\alpha = \frac{k-1}{2(k+1)}$$

and $c(k)$ is a constant depending only on k. (Theorem 5.3.1). Clearly the order of the upper bound in (5.1.3) is at most $O(n^{-\frac{1}{2}})$. It might be expected that the left hand side of (5.1.3) is, in fact, $O(n^{-\frac{1}{2}})$, but we have not been able to prove this.

The second main result of this Chapter concerns the multivariate Edgeworth expansion i.e. asymptotic expansion for the difference $(P_n - P)$. If $\beta_s < \infty$, then

$$(5.1.4) \qquad P_n(A) = \Phi(A) + \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} \int_A d\tilde{P}_j(-\Phi) + O\left(\frac{(\log n)^{\frac{1}{2}(k-1)}}{n^{\frac{1}{2}(s-2)}} \right),$$

A is any convex set, where the functions $\tilde{P}_j(-\Phi)$ are defined by 5.5.2. Here it is supposed that the c.f. of $Z^{(1)}$ satisfies Cramér's condition C. i.e. $\limsup_{|\underset{\sim}{t}| \to \infty} |f(\underset{\sim}{t})| < 1$. Once again it is to be expected that the $\log n$ factor in (5.1.4) is superfluous.

Note that if we suppose that $\beta_{s+1} < \infty$ then the remainder term in (5.1.4) is of the order $n^{-\frac{1}{2}(s-2)}$.

This Chapter consists of 9 sections. The second section contains preliminary material on general distributions in R_k. In section 3 we state the main results of this Chapter. Sections 4, 5, 6 and 7 develop the material for the proofs of the main theorems. The proofs of the main theorems are given in sections 8 and 9. The results on convex sets used in this Chapter will be found in Appendix A.

5.2. General remarks about distributions in R_k. Let $Z = (Z_1, \ldots, Z_k)$ be a k-dimensional random vector and let F be the probability distribution of Z i.e.

$$P[Z \in A] = \int_A dF = F(A) .$$

By the mean vector, or the expectation of Z we mean the vector $\underset{\sim}{m} = (m_1, \ldots, m_k)$ such that

$$m_i = E Z_i = \int x_i \, dF.$$

For all i, j $(1 \leq i, j \leq k)$ let

$$\sigma_{ij} = \int x_i x_j \, dF - E Z_i E Z_j$$
$$= \text{Cov}(Z_i, Z_j).$$

Then the matrix $\Sigma = \{\sigma_{ij}\}$ is called the variance covariance matrix of Z. It is assumed that the matrix Σ is positive definite.

The following notation will be employed for the moments of a random vector Z. Let

$$(5.2.1) \qquad \alpha_1^{s_1}\ \alpha_2^{s_2}\ \dots\ \alpha_k^{s_k} = \int x_1^{s_1}\ x_2^{s_2} \dots x_k^{s_k}\, dF$$

where $s_1, \dots, s_k$ are nonnegative integers. Here the product $\alpha_1^{s_1} \dots \alpha_k^{s_k}$ should be interpreted as a symbolic product so that we have

$$(t_1\alpha_1 + \dots + t_k\alpha_k)^s = \int (t_1x_1 + \dots + t_kx_k)^s\, dF.$$

Let

$$\beta_{js} = E\,|\,Z_j\,|^s = \int |x_j|^s\, dF, \quad \text{and}$$

$$(5.2.2) \qquad \beta_s = \sum_{j=1}^{k} \beta_{js}$$

$$\varrho_s = E\,|\,Z\,|^s = E(Z_1^2 + \dots + Z_k^2)^{s/2}.$$

Then it follows from Lyapounov's inequality that

$$\varrho_s \le k^{\frac{1}{2}(s-2)}\, \beta_s$$

If $Z^{(1)}$ and $Z^{(2)}$ are two independent random vectors with distributions F_1 and F_2 respectively, then the convolution of F_1 and F_2 (denoted by $F_1 * F_2$) is defined to be the probability distribution of the random vector $Z^{(1)} + Z^{(2)}$. Obviously $F_1 * F_2 = F_2 * F_1$. It is easy to see that

$$(F_1 * F_2)(E) = \int F_1(E - x)d\, F_2(x)$$

for each Borel set E. If F_1 and F_2 are both absolutely continuous with respect to Lebesgue measure i.e. there exist functions $f_1(\underset{\sim}{x})$, $f_2(\underset{\sim}{x})$ which are Lebesgue integrable and such that

$$F_1(A) = \int_A f_1(\underset{\sim}{x})\, d\underset{\sim}{x}\ ;\ F_2(A) = \int_A f_2(\underset{\sim}{x})\, d\underset{\sim}{x},$$

then the convolution $F_1 * F_2$ is also absolutely continuous, with density function $f_1 * f_2$, defined by

$$(5.2.3) \qquad (f_1 * f_2)(\underset{\sim}{x}) = \int f_1(\underset{\sim}{x}-\underset{\sim}{y})\, f_2(\underset{\sim}{y})\, d\underset{\sim}{y}.$$

As in the case of distributions on the real line, the chief tool in investigating the distribution of sums of random vectors is the characteristic function (c.f. in short). The characteristic function of the distribution F (or of the random vector Z) is defined as follows:

$$f(\underset{\sim}{t}) = f(t_1, \ldots, t_k) = \int e^{i\, \underset{\sim}{t}' \cdot \underset{\sim}{x}}\, d\, F(x)$$

where $\underset{\sim}{t}' \cdot \underset{\sim}{x}$ stands for the expression

$$t_1 x_1 + \ldots + t_k x_k .$$

Practically all the properties known to be true for real valued random variables have their analogues for random vectors (c.f. Cramér [10], Chapter 7). We just mention here that the c.f. determines the d.f. uniquely and the correspondence between c.f.s and d.f.s is bicontinuous i.e.

(i) $F_n \Rightarrow F$

is equivalent to

(ii) $f_n(\underset{\sim}{t}) \to f(\underset{\sim}{t})$ for each $\underset{\sim}{t}$, where f_n, f are the c.f.s of F_n and F respectively.

A very significant property of c.f.s is that the c.f. of $F_1 * F_2$ is $f_1(\underset{\sim}{t})\, f_2(\underset{\sim}{t})$ where f_1, f_2 are the c.f.s of F_1 and F_2 respectively i.e.

$$f_1(\underset{\sim}{t})\, f_2(\underset{\sim}{t}) = \int e^{i\, \underset{\sim}{t}' \cdot \underset{\sim}{x}}\, d\, F_1 * F_2 .$$

More generally the c.f. of $F_1 * F_2 * \ldots * F_n$ is $f_1(\underset{\sim}{t}) f_2(\underset{\sim}{t}) \ldots f_n(\underset{\sim}{t})$.

If we suppose that

$$\int |x_1|^{s_1} \dots |x_k|^{s_k} \, dF < \infty$$

then the c.f. $f(\underset{\sim}{t})$ has partial derivatives of all orders $(\nu_1, \dots, \nu_k)$ with $\nu_i \leq s_i$. It is easy to see that

$$\alpha_1^{s_1} \dots \alpha_k^{s_k} = \frac{1}{i^s} \left\{ \left(\frac{\partial}{\partial t_1} \right)^{s_1} \dots \left(\frac{\partial}{\partial t_k} \right)^{s_k} f \right\}_{\underset{\sim}{t}=0}$$

where $s = s_1 + s_2 + \dots + s_k$. Thus, if we suppose that $\beta_s < \infty$ then the following Taylor expansion holds:

$$f(\underset{\sim}{t}) = \sum_{j=1}^{s-1} \frac{i^j}{j!} (\alpha_1 t_1 + \dots + \alpha_k t_k)^j + \theta \frac{\varrho_s}{s!} |\underset{\sim}{t}|^s$$

where θ is some number such that $|\theta| \leq 1$.

We now proceed to define the semi-invariants or cumulants of a multidimensional distribution. Consider the function $\log f(\underset{\sim}{t})$. Here the logarithm is supposed to denote the principal branch. It is defined only for those $\underset{\sim}{t}$ for which $f(\underset{\sim}{t}) \neq 0$. It is also a continuous function of $\underset{\sim}{t}$ and takes the values zero at the origin. It may be verified that these conditions determine the branch uniquely.

Suppose that the absolute moment of order $(s_1, \dots, s_k)$ exists. Then the derivatives

∂t_1

$$\left(\frac{\partial}{\partial t_1}\right)^{s_1} \cdots \left(\frac{\partial}{\partial t_k}\right)^{s_k}$$

of both $f(\underset{\sim}{t})$ and $\log f(\underset{\sim}{t})$ exist and we define

$$(5.2.4) \quad \lambda_1^{s_1} \lambda_2^{s_2} \cdots \lambda_k^{s_k} = \frac{1}{i^s}\left\{\left(\frac{\partial}{\partial t_1}\right)^{s_1} \cdots \left(\frac{\partial}{\partial t_k}\right)^{s_k} \log f(\underset{\sim}{t})\right\}_{\underset{\sim}{t}=0}$$

where $s = s_1 + \ldots + s_k$. These are called the cumulants (or semi-invariants) of the distribution F. If we interpret the quantity

$$\lambda_1^{s_1} \cdots \lambda_k^{s_k}$$

as a symbolic product then it is clear that the following expansion holds:

$$(5.2.5) \quad \log f(\underset{\sim}{t}) = \sum_{j=1}^{s-1} \frac{i^j}{j!} (\lambda_1 t_1 + \ldots + \lambda_k t_k)^j + o\left(|\underset{\sim}{t}|^s\right)$$

There is a very useful relation between the c.f. $f(\underset{\sim}{t})$ and the c.f. of the real valued random variable Y defined as follows:

$$Y = (t_1 Z_1 + \ldots + t_k Z_k)/\,|\underset{\sim}{t}| \,.$$

Let $g(\tau)$ be the c.f. of Y i.e. $g(\tau) = E\, e^{i\tau Y}$. Then

$$g(\tau) = f\left(\frac{\tau}{|\underset{\sim}{t}|}\underset{\sim}{t}\right) = f\left(\frac{\tau}{|\underset{\sim}{t}|}\, t_1, \ldots, \frac{\tau}{|\underset{\sim}{t}|}\, t_k\right)$$

This is a very useful relation. For instance suppose that $\varrho_s = E\ |\underset{\sim}{Z}|^s < \infty$. Then $E\ |Y|^s \leq E\ |\underset{\sim}{Z}|^s < \infty$, so that the random variable Y has cumulants upto the sth order, say $\nu_1, \ldots, \nu_{s-1}$. Since

$$\frac{d}{d\tau}\ g(\tau) = \frac{1}{|t|}\left(t_1 \frac{\partial}{\partial u_1} + \ldots + t_k \frac{\partial}{\partial u_k}\right) f(\underset{\sim}{u})$$

at the point $\underset{\sim}{u} = \tau\ \underset{\sim}{t}/|\underset{\sim}{t}|$, it follows that

$$(5.2.6)\quad \nu_j = \frac{1}{i^j}\left\{\frac{d^j}{d\tau^j}\left\{\log g(\tau)\right\}\right\}_{\tau = 0}$$

$$= \frac{1}{i^j}\ \frac{1}{|\underset{\sim}{t}|^j}\left\{\left(t_1 \frac{\partial}{\partial u_1} + \ldots + t_k \frac{\partial}{\partial u_k}\right) f(u)\right\}_{\underset{\sim}{u} = 0}$$

$$= \frac{1}{i^j}(t_1 \lambda_1 + \ldots + t_k \lambda_k)^j\ \frac{1}{|\underset{\sim}{t}|^j}$$

Thus the expansion (5.2.5) could be deduced from the corresponding expansion of $g(\tau)$ viz.

$$\log g(\tau) = \sum_{j=1}^{s} \frac{(i\tau)^j}{j!} + o\ (|\tau|^s)$$

In our investigations it is important to know the behaviour of the c.f. $f(\underset{\sim}{t})$ as $|\underset{\sim}{t}| \to \infty$. The following results in this direction are well known (c.f. Esseen [15] Chapter 1). Suppose F is absolutely continuous with respect to Lebesgue measure i.e. there

exists a density $p(x)$ such that

$$F(A) = \int_A p(\underset{\sim}{x})\, d\underset{\sim}{x}$$

Then from the multidimensional analogue of the Riemann-Lebesgue theorem it follows that

$$\lim_{|\underset{\sim}{t}| \to \infty} |f(\underset{\sim}{t})| = 0. \tag{5.2.7}$$

In the general case let

$$F = \alpha F_1 + \beta F_2 + \gamma F_3, \tag{5.2.8}$$

where $\alpha, \beta, \gamma \geq 0$ and $\alpha + \beta + \gamma = 1$. In (5.2.8) F_1, F_2, F_3 are three distributions such that (1) F_1 is absolutely continuous with respect to the Lebesgue measure, (2) F_2 is continuous but singular and (3) F_3 is purely discontinuous. If f_1, f_2, f_3 denote their c.f.s then

$$f = \alpha f_1 + \beta f_2 + \gamma f_3.$$

Thus if $\alpha > 0$ i.e. F has an absolutely continuous component then from (5.2.7)

$$\limsup_{|\underset{\sim}{t}| \to \infty} |f(\underset{\sim}{t})| \leq 1 - \alpha < 1.$$

In case the distribution is purely singular i.e. $\alpha = \gamma = 0$, then lim sup $|f(\underset{\sim}{t})|$ can be zero, less than one or equal to one. (cf. Esseen [15] Chapter 1). If F is purely discrete i.e. $\alpha = \beta = 0$, then it is not difficult to show that always we have

$$(5.2.9) \qquad \limsup_{|t| \to \infty} |f(\underset{\sim}{t})| = 1.$$

A simple class of distributions which satisfy (5.2.9) are those for which

$$|f(t_o)| = 1$$

for some finite point t_o. These distributions are known as lattice distributions and they will be discussed in the next Chapter.

We conclude this section by introducing the normal distribution. Let Σ be a nonsingular symmetric and positive definite matrix and let $\underset{\sim}{m}$ be an arbitrary vector. Then the normal distribution $N(\underset{\sim}{m}, \Sigma)$ with mean $\underset{\sim}{m}$ and variance covariance matrix Σ is defined to be the distribution with the density

$$\frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \exp\left\{ -\tfrac{1}{2}(\underset{\sim}{x} - \underset{\sim}{m})' \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{m}) \right\}$$

The c.f. of $N(\underset{\sim}{m}, \Sigma)$ is

$$\exp\left\{ i\, \underset{\sim}{t}' \cdot \underset{\sim}{m} - \frac{1}{2}\, \underset{\sim}{t}' \Sigma\, \underset{\sim}{t} \right\}$$

14

The distribution $N(0, I)$, where I is the identity matrix, will be denoted by Φ i.e.

$$\Phi(A) = \frac{1}{(2\pi)^{k/2}} \int_A \exp\left(-\frac{1}{2} \sum_{j=1}^{k} x_j^2\right) dx_1 \ldots dx_k .$$

and the distribution function by $\Phi(x)$, i.e.

$$\Phi(x) = (2\pi)^{-\frac{1}{2}k} \int_{-\infty < \underset{\sim}{y} \leq \underset{\sim}{x}} \exp\left(-\frac{1}{2} |\underset{\sim}{y}|^2\right) d\underset{\sim}{y}.$$

5.3. The main Theorems. Throughout the rest of the Chapter we adopt the following notation. $Z^{(1)}, Z^{(2)}, \ldots, Z^{(n)}, \ldots$, denote a sequence of independent and identically distributed random vectors in R_k. We suppose that the following conditions are fulfilled:

1) the mean value of each component is zero, i.e. $E\, Z_j^{(1)} = 0$ for $j = 1, 2, \ldots k$;

2) the variance covariance matrix of $Z^{(1)}$ is the identity matrix i.e. $E\, Z_r^{(1)} Z_s^{(1)} = 0$ for $r \neq s$ and $E\,[Z_r^{(1)}]^2 = 1$. It is pointed out later that the assumption 1 and 2 involve no loss of generality.

All moments and cumulants occuring in the sequel refer to the random vector $Z^{(1)}$. Let

$$Y^{(n)} = \frac{Z^{(1)} + \ldots + Z^{(n)}}{\sqrt{n}}$$

and let P_n be the probability distribution of $Y^{(n)}$ i.e.

15

$$P_n(A) = \int_A d\, P_n = P\left[\, Y^{(n)} \varepsilon A \,\right]$$

and we will denote by $F_n(\underset{\sim}{x})$ the distribution function of $Y^{(n)}$ i.e. for $\underset{\sim}{x} = (x_1, \ldots, x_k)$

$$F_n(\underset{\sim}{x}) = P\left[\, Y_j^{(n)} \leq x_j \quad \text{for} \quad j = 1, 2, \ldots, k. \right].$$

Let $\underset{\sim}{\Phi}(\underset{\sim}{x})$ denote the normal distribution with mean zero and variance covariance matrix I and $\underset{\sim}{\Phi}(A) = \int_A d\, \underset{\sim}{\Phi}(\underset{\sim}{x})$. We may then state the main theorems as follows.

Theorem 5.3.1. Suppose $\beta_4 < \infty$. Then

$$(5.3.1) \qquad \sup_{A \varepsilon \mathcal{C}} |P_n(A) - \underset{\sim}{\Phi}(A)| \leq c(k)\, \beta_4^{3/2}\, \frac{(\log n)^{\alpha}}{\sqrt{n}}$$

where $\alpha = (k-1)/2(k+1)$, $c(k)$ is a constant depending only on k and $\mathcal{C}$ is the class of all convex subsets of R_k.

Theorem 5.3.2. Suppose the c.f $f(\underset{\sim}{t})$ of $Z^{(1)}$ satisfies condition (C), i.e.

$$(5.3.2) \qquad \limsup_{|\underset{\sim}{t}| \to \infty} |f(\underset{\sim}{t})| < 1.$$

If $\beta_s < \infty$, then

$$(5.3.3)\qquad P_n(A) = \Phi(A) + \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} \int_A d\tilde{P}_j(-\Phi) + O\left(\frac{(\log n)^{\frac{1}{2}(k-1)}}{n^{\frac{1}{2}(s-2)}}\right)$$

uniformly for $A \in \mathcal{C}$, where the functions $\tilde{P}_j(-\Phi)$ are defined by (5.5.1).

The method employed in the proof of these theorems is a convolution method based on the work of Esseen. In particular we utilize his results on the value distribution of $|f(t)|$. The proofs of these two theorems forms the content of the rest of the Chapter.

Remarks 1. Since the class $\mathcal{C}$ of all convex subsets of R_k is invariant under any linear transformation of the space R_k into itself, it is clear that there is no loss in generality in supposing, as we have done, that the mean vector of $Z^{(1)}$ is zero and that the variance covariance matrix of $Z^{(1)}$ is the identity matrix. In the general case, if $\underset{\sim}{m}$ is the mean and Σ is the variance covariance matrix of $Z^{(1)}$, we have from Theorem 5.3.1 that

$$(5.3.4)\qquad \sup_{A \in \mathcal{C}} |P_n(A) - \Phi(A)| \leq c(k)\, \beta_4^{3/2}\, \gamma^{3/2}\, \frac{(\log n)^{\alpha}}{\sqrt{n}}$$

where $\alpha = (k-1)/2(k+1)$ and γ is given by

$$\gamma = \left[\min_j b_j\right]^{-1},$$

17

$b_1, \ldots b_k$ being the eigen-values of Σ; here Φ denotes the normal distribution with mean zero and variance covariance matrix Σ. Obviously Theorem 5.3.2 can also be restated in the same way for the general case.

2. A bound similar to (5.3.1) could be obtained if we suppose that only third moments are finite. For instance the following inequality holds:

$$\sup_{A \in \mathcal{C}} |P_n(A) - \Phi(A)| \leq c(k)\, \beta_3 \, \frac{(\log n)^{\frac{1}{2}(k-1)}}{\sqrt{n}}.$$

The proof of this inequality is entirely analogous to that of Theorem 5.3.1. We have assumed the existance of fourth moments only because this enables us to reduce the power of (log n) considerably. As a matter of fact it appears that the log n factor is entirely superfluous. Some evidence to this effect is afforded by Theorem 5.3.2. For if condition (C) is fulfilled and $\beta_4 < \infty$ then by (5.3.3)

$$\sup_{A \in \mathcal{C}} |P_n(A) - \Phi(A)| = O(n^{-\frac{1}{2}}).$$

A similar result holds for the case when $Z^{(1)}$ is a lattice variable; this follows from the asymptotic expansion for the lattice case given in the next Chapter.

3. It should be noted that if $\beta_{s+1} < \infty$ then the order of the remainder in Theorem 5.3.2 could be improved. Indeed, by changing s to s+1 in Theorem 5.3.2 we have

$$(5.3.5) \qquad P_n(A) = \Phi(A) + \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} \int_A d\tilde{P}_j \left(-\Phi\right) + o\left(n^{-\frac{1}{2}(s-2)}\right).$$

In particular if $Z^{(1)}$ has moments of all orders then

$$P_n(A) = \Phi(A) + \sum_{j=1}^{s} n^{-\frac{1}{2}j} \int_A d\,\tilde{P}_j \left(-\Phi\right) + o\left(n^{-\frac{1}{2}(s+1)}\right)$$

for every integer $s \geq 0$.

4. The asymptotic expansion for the distribution function $F_n(\underset{\sim}{x})$ of $Y^{(n)}$ can be deduced from 5.3.3, as a special case. In fact if $\beta_s < \infty$ then

$$(5.3.6) \quad F_n(\underset{\sim}{x}) = \Phi(\underset{\sim}{x}) + \sum_{j=0}^{s-3} n^{-\frac{1}{2}j} \tilde{P}_j(-\Phi) + o\left\{\frac{(\log n)^{\frac{1}{2}(k-1)}}{n^{\frac{1}{2}(s-2)}}\right\}$$

uniformly in the vector $\underset{\sim}{x}$. Again if we take

$$A = [\, \underset{\sim}{x} : \; | \underset{\sim}{x} | \leq a \,]$$

19

then it follows easily from Theorem 5.3.2 that

$$(5.3.7) \qquad P_n(A) = P[\chi_k^2 \leq a^2] + O\left(\frac{1}{n}\right)$$

where χ_k^2 is the chi-square distribution with k-degrees of freedom and the estimation in (5.3.7) holding uniformly in a.

5.4. Expansions of the characteristic function of $Y^{(n)}$. Let $f(\underset{\sim}{t}) = f(t_1, \ldots, t_k)$ be the c.f. of the vector $Z^{(1)}$ and $f_n(\underset{\sim}{t})$ be that of $Y^{(n)}$. Then obviously

$$f_n(\underset{\sim}{t}) = [f(\underset{\sim}{t}\, n^{-\frac{1}{2}})]^n$$

By elementary calculations it can be proved that

$$(5.4.1) \qquad \lim_{n \to \infty} f_n(\underset{\sim}{t}) = \exp\left(-\tfrac{1}{2} |\underset{\sim}{t}|^2\right)$$

for all $\underset{\sim}{t}$. The central limit theorem in R_k is a consequence of (5.4.1) and the continuity of the correspondence between c.f s and d.f.s. Our object in this section is to obtain an asymptotic expansion of the difference

$$(5.4.2) \qquad f_n(\underset{\sim}{t}) - \exp\left(-\tfrac{1}{2} |\underset{\sim}{t}|^2\right)$$

with estimates of the remainder. Such an expansion for the case

19 $k = 1$ is indeed classical and has been given by Charlier and Edgeworth. The estimates of the remainder in the case $k = 1$ have been given by Cramér [10] (cf. also Esseen [15]) Actually, the expansion of the function (5.4.2) involves only elementary computations with the Taylor expansion of log $f_n(\underset{\sim}{t})$. We shall, however, follow an indirect method of deducing it from the corresponding result in R_1.

Let us suppose that $\beta_s < \infty$. Let $\lambda_1^{s_1}\ \lambda_2^{s_2}, \ldots, \lambda_k^{s_k}$ denote the cumulants of $Z^{(1)}$. Then it follows from (5.2.5) that the following expansion holds:

$$\log\ f(\underset{\sim}{t}) = -\tfrac{1}{2}\,|\underset{\sim}{t}|^2 + \sum_{j=3}^{s} \frac{i^j}{j!}\,(\lambda_1 t_1 + \ldots + \lambda_k t_k)^j + o\,(|\underset{\sim}{t}|^s).$$

Consequently,

$$\log\ f_n(\underset{\sim}{t}) = n \log\ f(\underset{\sim}{t}\, n^{-\frac{1}{2}}) = -\tfrac{1}{2}\,|\underset{\sim}{t}|^2 + \sum_{j=3}^{s} \frac{i^j}{j!}\,(\lambda_1 t_1 + \ldots + \lambda_k t_k)^j\ n^{-\frac{1}{2}(j-2)} + o\,(|\underset{\sim}{t}|^s\, n^{-\frac{1}{2}(s-2)}). \tag{5.4.3}$$

Let the polynomials $P_j(\underset{\sim}{w})$ be defined by the identity :

$$\exp\left\{\sum_{j=3}^{\infty}(\lambda_1 w_1 + \cdots + \lambda_k w_k)^j \frac{n^{-\frac{1}{2}(j-2)}}{j!}\right\} = 1 + \sum_{j=1}^{\infty} n^{-\frac{1}{2}j}\,\tilde{P}_j(\underset{\sim}{w}). \tag{5.4.4}$$

The first two polynomials are easily calculated:

$$P_1(w) = \frac{1}{6}(\lambda_1 w_1 + \cdots + \lambda_k w_k)^3 \tag{5.4.5}$$

$$P_2(w) = \frac{1}{24}(\lambda_1 w_1 + \cdots + \lambda_k w_k)^4 + \frac{1}{72}\left\{(\lambda_1 w_1 + \cdots + \lambda_k w_k)^3\right\}^2 \tag{5.4.6}$$

It is convenient to define $P_0(w) \equiv 1$.

From (5.4.3) and (5.4.4) it is evident that

$$f_n(\underset{\sim}{t}) = \exp\left(-\frac{1}{2}|\underset{\sim}{t}|^2\right)\left[1 + \sum_{j=1}^{s-2} n^{-\frac{1}{2}j}\,\tilde{P}_j(i\underset{\sim}{t})\right] + O\left(n^{-\frac{1}{2}(s-2)}\right)]$$

A more precise estimate of the remainder is given by the following theorem.

Theorem 5.4.1. For $|\underset{\sim}{t}| < \sqrt{n}\,\frac{\varrho_s^{-3/s}}{8s}$, the following inequality holds:

$$(5.4.7)\qquad \left| f_n(\underset{\sim}{t}) - \exp\left(-\tfrac{1}{2}|\underset{\sim}{t}|^2\right) \sum_{j=0}^{s-3} n^{-\frac{1}{2}j} \widetilde{P}_j(i\underset{\sim}{t}) \right|$$

$$\leq c_1(s)\ \frac{\varrho_s^{3(s-2)/s}}{n^{\frac{1}{2}(s-2)}} \left\{ |\underset{\sim}{t}|^{3(s-2)} + |\underset{\sim}{t}|^s \right\} \exp\left(-\tfrac{1}{4}|t|^2\right).$$

where $c_1(s)$ is a constant depending only on s and ϱ_s is defined by (5.2.2). For $\underset{\sim}{t}$ belonging to the same range, the following inequality also holds:

$$(5.4.8)\qquad \left| f_n(\underset{\sim}{t}) - \exp\left(-\tfrac{1}{2}|\underset{\sim}{t}|^2\right) \left\{ \sum_{j=0}^{s-2} n^{-\frac{1}{2}j} \widetilde{P}_j(i\underset{\sim}{t}) \right\} \right|$$

$$\leq \delta(n)\ \frac{c_2(s)\ \varrho_s^{3(s-2)/s}}{n^{\frac{1}{2}(s-2)}} \left\{ |\underset{\sim}{t}|^s + |\underset{\sim}{t}|^{3(s-2)} \right\} e^{-|t|^2/4}$$

where $c_2(s)$ depends only on s and

$$\lim_{n \to \infty} \delta(n) = 0 .$$

Proof. Consider the real valued random variable ξ defined by

$$\xi = (t_1 Z_1^{(1)} + \ldots + t_k Z_k^{(1)})/|\underset{\sim}{t}| .$$

Evidently $E\xi = 0$, $E\xi^2 = 1$, and

$$(5.4.9)\qquad E|\xi|^s \leq E|Z^{(1)}|^s = \varrho_s < \infty .$$

Let $g(\tau)$ for real τ, be the c.f. of ξ i.e.

$$g(\tau) = E\, e^{i\tau\xi} = f\left(\frac{\tau}{|\underset{\sim}{t}|}\ \underset{\sim}{t}\right)$$

22 Also

$$(5.4.10) \qquad g_n(\tau) = [g(\tau\, n^{-\frac{1}{2}})]^n = f_n\Big(\frac{\tau}{|\underset{\sim}{t}|}\ \underset{\sim}{t}\Big)$$

The theorem is then deduced from the asymptotic expansion for $g_n(\tau)$. It is known that (cf. Cramér [10], p. 71); for $|\tau| < \sqrt{n}\left\{8s\,(E\,|\xi|^s)^{3/s}\right\}^{-1}$,

$$(5.4.11) \qquad \Big|g_n(\tau) - e^{-\frac{1}{2}\tau^2}\Big\{\sum_{j=0}^{s-3} n^{-\frac{1}{2}j}\, P_j(i\tau)\Big\}\Big| \le \frac{c_1(s)\left\{8s\,(E|\xi|^s)^{3/s}\right\}^{s-2}}{n^{\frac{1}{2}(s-2)}}\ \underbrace{\left[|\tau|^s + |\tau|^{3(s-2)}\right]}_{\exp\,(-|\tau|^2/4)}$$

Here the polynomial $P_j(i\tau)$ is constructed from the cumulants of $g(\tau)$ from an identity similar to (5.4.4). From formula (5.2.6) and (5.4.4) it is not difficult to deduce that

$$(5.4.12) \qquad P_j(i\tau) = \tilde{P}_j\Big(i\,\frac{\tau}{|\underset{\sim}{t}|}\ \underset{\sim}{t}\Big)$$

The inequality (5.4.7) is then an immediate consequence of (5.4.9) - (5.4.12) if we substitute $\tau = |\underset{\sim}{t}|$ in (5.4.11). The second inequality (5.4.8) can be deduced in the same way from the corresponding expansion for $g_n(\tau)$ (cf. Esseen [15], p. 44).

23 This completes the proof.

5.5. The functions $\tilde{P}_j(-\overline{\underline{\phi}})$. In the last section the polynomials $\tilde{P}_j(\underset{\sim}{w})$ in the vector $\underset{\sim}{w} = (w_1, \ldots, w_k)$ were defined by the formal identity (5.4.4). Let D stand for the operator

$$D = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_k}\right)$$

Let $\tilde{P}_j(-D)$ denote the operator obtained when we substitute the operator $(-D)$ for the vector $\underset{\sim}{w}$ in the polynomial $\tilde{P}_j(\underset{\sim}{w})$. Then we define

$$(5.5.1) \qquad \tilde{P}_j(-\overline{\underline{\phi}}) = P_j(-D)\,\overline{\underline{\phi}}$$

where the function $\overline{\underline{\phi}}$ is defined in section 5.2. Thus, for instance,

$$\tilde{P}_1(-\overline{\underline{\phi}}) = -\frac{1}{6}\left(\lambda_1 \frac{\partial}{\partial x_1} + \cdots + \lambda_k \frac{\partial}{\partial x_k}\right)^3 \overline{\underline{\phi}}(\underset{\sim}{x})$$

which on simplification reduces to

$$(5.5.2) \qquad \tilde{P}_1(-\overline{\underline{\phi}}) = \frac{1}{6}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}\left(\sum \alpha_j x_j\right)^3 e^{-\frac{1}{2}|\underset{\sim}{x}|^2}\,dx_1\ldots dx_k$$

$$-\frac{1}{2}\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}\left(\sum \alpha_j^2\right)\left(\sum \alpha_j x_j\right) e^{-\frac{1}{2}|\underset{\sim}{x}|^2}\,dx_1\ldots dx_k$$

where the α's are the moments of $Z^{(1)}$ defined by (5.2.1).

24 It is clear from the definition that the $\tilde{P}_j(-\Phi)$ are all functions of bounded variation. They are also absolutely continuous with respect to Lebesgue measure in R_k, for

$$\frac{\partial^k}{\partial x_1 \cdots \partial x_k} \tilde{P}_j(-\Phi) = \tilde{P}_j(-D) \frac{\partial^k \Phi}{\partial x_1 \cdots \partial x_k}$$
$$= \tilde{P}_j(-D)\, \varphi(\underset{\sim}{x})$$

where $\varphi(x)$ is the density of Φ. We define

$$(5.5.3) \qquad \tilde{P}_j(-\varphi) = \tilde{P}_j(-D)\, \varphi(\underset{\sim}{x}).$$

The Fourier transforms of $\tilde{P}_j(-\Phi)$ are given by the following lemma.

Lemma 5.5.1. For each integer $j \geq 0$,

$$(5.5.4) \qquad \tilde{P}_j(it)\, e^{-\frac{1}{2}|\underset{\sim}{t}|^2} = \int e^{i\underset{\sim}{t}'\cdot\underset{\sim}{x}}\, d\, \tilde{P}_j(-\Phi)$$
$$= \int e^{i\underset{\sim}{t}'\cdot\underset{\sim}{x}}\, \tilde{P}_j(-\varphi)\, d\underset{\sim}{x}$$

Proof. It follows from the Fourier inversion theorem that

$$\varphi(\underset{\sim}{x}) = \int e^{i\,\underset{\sim}{t}'\cdot\underset{\sim}{x} - \frac{1}{2}|\underset{\sim}{t}|^2}\, d\underset{\sim}{t}.$$

Consequently

$$\frac{\partial^r}{\partial x_j^r}\varphi = \int (it_j)^r\, e^{i\underset{\sim}{t}'\cdot\underset{\sim}{x} - \frac{1}{2}|\underset{\sim}{t}|^2}\, d\underset{\sim}{t}$$

25 or more generally

$$\frac{\partial^{r_1+r_2+\ldots+r_k}}{\partial x_1^{r_1} \ldots \partial x_k^{r_k}} \varphi(x) = \int (i t_1)^{r_1} \ldots (i t_k)^{r_k} \exp(i \underset{\sim}{t}' . \underset{\sim}{x} - \tfrac{1}{2}|\underset{\sim}{t}|^2) d\underset{\sim}{t}$$

Thus if $Q(\underset{\sim}{w})$ is any polynomial in the vector $\underset{\sim}{w} = (w_1, \ldots, w_k)$ then

$$Q(D)\, \varphi(\underset{\sim}{x}) = \int Q(i\underset{\sim}{t}) \exp(i\underset{\sim}{t}' . \underset{\sim}{x} - \tfrac{1}{2}|\underset{\sim}{t}|^2) d\underset{\sim}{t}$$

where D stands for the operator $D = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_k})$. Since $Q(D)\, \varphi(\underset{\sim}{x})$ is absolutely integrable it follows from the inversion theorem that

$$Q(i\underset{\sim}{t})\, e^{-\frac{1}{2}|\underset{\sim}{t}|^2} = \int Q(D)\, \varphi(\underset{\sim}{x}) d\underset{\sim}{x}$$

The lemma follows by taking $Q(\underset{\sim}{w}) = \tilde{P}_j(\underset{\sim}{w})$.

Remark. We are assuming that the variance covariance matrix of $Z^{(1)}$ is the identity matrix. Consequently the polynomials $\tilde{P}_j(\underset{\sim}{w})$ or the functions $\tilde{P}_j(-\Phi)$ are defined only for this case. However, this definition may be easily extended for the general case when the variance covariance matrix of $Z^{(1)}$, say Σ, is $\neq 1$. In this case the polynomials $\tilde{P}_j(\underset{\sim}{w})$ will be defined again by the identity (5.4.4) and $\tilde{P}_j(-\Phi)$ by the relation 5.5.1 where now it is natural to suppose that Φ is the normal d.f. with mean zero and covariance matrix Σ. With this definition it is not difficult to see that analogues of Theorem 5.4.1. and lemma 5.5.1

26 hold in this case also. The reason for pointing out these definitions here is that it will necessary to consider the case $\Sigma \neq I$ when we discuss lattice vectors in the next Chapter.

5.6. Some lemmas concerning $f_n(t)$. In this section we present two theorems which are needed later. Both these results are in essence due to Esseen ([15], Chapter 7).

Theorem 5.6.1. Let S_ϱ denote an arbitrary sphere of radius ϱ, where

$$\varrho = \frac{1}{6(1 + 2^{1/3})(k\beta_4)^{3/4}} \tag{5.6.1}$$

Then there exists a constant $c(k)$ depending only on k such that

$$\int_{S_\varrho} |f(\underset{\sim}{t})|^n \, d\underset{\sim}{t} \leq \frac{c(k)}{n^{k/2}} . \tag{5.6.2}$$

Theorem 5.6.2. There exists a constant $c(k)$ depending only on k such that for every $\beta > \varrho$, $\alpha > k$, we have

$$\int_{|\underset{\sim}{t}| > \beta} \frac{|f(t)|^n}{|\underset{\sim}{t}|^\alpha} \, d\underset{\sim}{t} \leq \frac{c(k)}{n^{k/2}} \left(\frac{1}{\varrho}\right)^{\alpha-k+1} \tag{5.6.2}$$

where ϱ is given by (5.6.1).

Proof. Let K_δ denote an arbitrary cube with sides parallel to the axes and edge length δ

27

(5.6.3) $$\delta = \frac{2\varrho}{\sqrt{k}} ,$$

where ϱ is given by (5.6.1). Let D_o be a cube (all cubes considered here have sides parallel to the axes) inscribed in a sphere with centre at the origin and radius β. Then the edge length of D_o is

$$b = \frac{2\beta}{\sqrt{k}} .$$

Let $D_1, D_2, \ldots$ be a sequence of cubes homothetic to D_o such that $D_o \subset D_1 \subset D_2 \ldots$, where the edge length of D_m is given by

$$2(b + m\delta).$$

Then it is clear that

$$(5.6.4) \qquad \int_{|\underset{\sim}{t}| > \beta} \frac{|f(t)|^n}{|\underset{\sim}{t}|^\alpha} d\underset{\sim}{t} \leq \int_{R_k - D_o} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n d\underset{\sim}{t}$$

$$\leq \sum_{j=1}^{\infty} \int_{D_j - D_{j-1}} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n d\underset{\sim}{t}$$

Also

$$(5.6.5) \qquad \int_{D_j - D_{j-1}} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n d\underset{\sim}{t} \leq N_j \operatorname*{Max}_{K_\delta} \int_{K_\delta} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n d\underset{\sim}{t}$$

where N_j is the number of cubes K_δ required to cover $D_j - D_{j-1}$

28

and the supremum in the right hand side of the relation (5.6.5) is taken over all cubes $K_\delta \subset D_j - D_{j-1}$. It is not difficult to show that

$$(5.6.6) \qquad N_j \leq 2^k (b + j\delta)^{k-1}$$

for $j = 1, 2, \ldots$. Also, for $\underset{\sim}{t} \in D_j - D_{j-1}$

$$(5.6.7) \qquad |\underset{\sim}{t}| > b + (j-1)\delta .$$

Combining (5.6.5), (5.6.6) and (5.6.7) we deduce that

$$(5.6.8) \quad \int_{D_j - D_{j-1}} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n \, dt \leq \frac{2^k (b + j\delta)^{k-1}}{(b+(j-1)\delta)^\alpha} \times \sup_{K_\delta} \int_{K_\delta} |f(\underset{\sim}{t})|^n \, d\underset{\sim}{t}$$

From (5.6.8) and Theorem 5.6.1 it follows that

$$(5.6.9) \quad \int_{D_j - D_{j-1}} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n d\underset{\sim}{t} \leq \frac{c_1(k)}{n^{k/2}} \cdot \frac{1}{(b+\delta(j-1))^{\alpha-k+1}}$$

Since $b > \delta$ it follows from (5.6.9) that

$$(5.6.10) \quad \sum_{j=1}^{\infty} \int_{D_j - D_{j-1}} |\underset{\sim}{t}|^{-\alpha} |f(\underset{\sim}{t})|^n \, d\underset{\sim}{t} \leq \frac{c_1(k)}{n^{k/2}} \left(\frac{1}{\delta}\right)^{\alpha-k+1} \sum_{j=1}^{\infty} j^{-(\alpha-k+1)}$$

Since $\alpha > k$,

$$\sum j^{-(\alpha-k+1)} < \infty .$$

The assertion of the theorem is then an immediate consequence of (5.6.3), (5.6.4) and (5.6.10). This completes the proof.

5.7. Some auxiliary functions. For any set A, let $\chi(x;A)$ denote its characteristic function i.e

$$\chi(\underset{\sim}{x}\ ;\ A) = \begin{cases} 1 & \text{if } \underset{\sim}{x} \in A \\ 0 & \text{otherwise} \end{cases}$$

Let $S(\underset{\sim}{x}, a)$ denote the sphere with centre at $\underset{\sim}{x}$ and radius a. For any set A and any $\alpha > 0$, let the sets A^{α}, $A^{-\alpha}$ be defined as follows :

$$A^{\alpha} = [\ \cup S(\underset{\sim}{x}, \alpha)\ ,\ \underset{\sim}{x} \in A\ ]$$

$$A^{-\alpha} = [\ \underset{\sim}{x} :\ S(\underset{\sim}{x}, \alpha) \subset A\ ]$$

It is clear that $A^{-\alpha}$ has non empty interior if and only if its in-radius $r(A) > \alpha$. (For definition of in-radius see Appendix A).

Let the functions $Q_{\varepsilon}(\underset{\sim}{x})$ and $M(\underset{\sim}{x}, A)$ be defined as follows:

$$(5.7.1) \qquad Q_{\varepsilon}(\underset{\sim}{x}) = c_{\varepsilon}\, \chi(\underset{\sim}{x}\ ;\ S(0, \varepsilon))$$

where the constant c_{ε} is determined by

$$(5.7.2) \qquad \int Q_{\varepsilon}(\underset{\sim}{x})\, dx = 1$$

It is easy to show that

30 (5.7.3) $$c_\varepsilon = \frac{\Gamma(1 + k/2)}{\pi^{k/2}\, \varepsilon^k} .$$

For any set A with positive Lebesgue measure let

(5.7.4) $$M(\underset{\sim}{x}; A) = \chi(\underset{\sim}{x}; A) * Q_\varepsilon^{*2}$$

where the convolution is defined by (5.2.3).

Let $q_\varepsilon(\underset{\sim}{t})$ be the Fourier transform of $Q_\varepsilon(\underset{\sim}{x})$. i.e.

$$q_\varepsilon(\underset{\sim}{t}) = \int e^{i\, \underset{\sim}{t}' \cdot \underset{\sim}{x}} Q_\varepsilon(\underset{\sim}{x})\, d\underset{\sim}{x}$$

From (5.7.2) it follows that, for all $\underset{\sim}{t}$

(5.7.5) $$|q_\varepsilon(\underset{\sim}{t})| \leq 1$$

Then it can be shown (cf. Bochner [6] section 43) that

(5.7.6) $$q_\varepsilon(\underset{\sim}{t}) = \Gamma(1 + k/2) \left(\frac{2}{\varepsilon |\underset{\sim}{t}|}\right)^{k/2} J_{k/2}(\varepsilon\, |\underset{\sim}{t}|)$$

where $J_{k/2}(z)$ is the Bessel function of order $k/2$.
From the property of Bessel functions it is easily shown (cf. Watson [40], p.199) that

(5.7.7) $$|J_{k/2}(z)| \leq \frac{c}{\sqrt{z}}$$

for all positive z where c is a constant. It follows from (5.7.6) and (5.7.7) that

$$(5.7.8) \qquad |q_\varepsilon(\underset{\sim}{t})| \leq \frac{c(k)}{(\varepsilon\,|\underset{\sim}{t}|)^{(k+1)/2}}$$

where $c(k)$ is a constant depending only on k.

Let $q(\underset{\sim}{t}\,;A)$ and $m(\underset{\sim}{t};A)$ be the Fourier transforms of $\lambda(\underset{\sim}{x}\,;A)$ and $M(\underset{\sim}{x}\,;A)$ respectively. (We suppose that A has positive Lebesgue measure). Then it follows from the definition (5.7.4) of $M(\underset{\sim}{x}\,;A)$ that

$$(5.7.9) \qquad m(\underset{\sim}{t}\,;A) = q(\underset{\sim}{t}\,;A)\, q_\varepsilon^2(\underset{\sim}{t}).$$

It may noted here that $q_\varepsilon(\underset{\sim}{t})$ being the Fourier transform of a bounded integrable function is square integrable, so that

$$(5.7.10) \qquad \int |q_\varepsilon(\underset{\sim}{t})|^2\, d\underset{\sim}{t} < \infty \text{ and } \int |m(\underset{\sim}{t}\,;A)|\, d\underset{\sim}{t} < \infty .$$

Lemma 5.6.1. For any set A with positive Lebesgue measure and $\varepsilon > 0$, we have

$$(5.7.11) \qquad \lambda(\underset{\sim}{x}\,;A^{-4\varepsilon}) \leq M(\underset{\sim}{x}\,;A^{-2\varepsilon}) \leq \lambda(\underset{\sim}{x}\,;A)$$

$$(5.7.12) \qquad \lambda(\underset{\sim}{x}\,;A) \leq M(\underset{\sim}{x}\,;A^{2\varepsilon}) \leq \lambda(\underset{\sim}{x}\,;A^{4\varepsilon})$$

32 Proof. From the definition (5.7.1) of $Q_\varepsilon(\underset{\sim}{x})$ it is easily shown that

1) $\int Q_\varepsilon^{*2}(\underset{\sim}{x})\, d\underset{\sim}{x} = 1,$

2) $0 \leq Q_\varepsilon^{*2}(\underset{\sim}{x}) \leq 1$ for all $\underset{\sim}{x}$,

3) $Q_\varepsilon^{*2}(\underset{\sim}{x}) = 0$ for $|\underset{\sim}{x}| > 2\varepsilon$.

Let B be any Borel set. Then the definition (5.7.4) and the property(3) above together imply that

$$(5.7.13)\quad M(\underset{\sim}{x};\, B) = \int_{|\underset{\sim}{y}| \leq 2\varepsilon} \lambda(\underset{\sim}{x}-\underset{\sim}{y};\, B)\, Q_\varepsilon^{*2}(\underset{\sim}{y})\, d\underset{\sim}{y}.$$

Suppose $\underset{\sim}{x} \in B^{-2\varepsilon}$. Then the definition of $B^{-2\varepsilon}$ implies that

$$\underset{\sim}{x} - \underset{\sim}{y} \in B, \quad \text{whenever} \quad |\underset{\sim}{y}| \leq 2\varepsilon.$$

Thus it follows from (5.6.13), that $M(\underset{\sim}{x}, B) = 1$ whenever $\underset{\sim}{x} \in B^{-2\varepsilon}$ or

$$(5.7.14)\quad M(\underset{\sim}{x};\, B) \geq \lambda(\underset{\sim}{x};\, B^{-2\varepsilon}).$$

Atain if $\underset{\sim}{x} \notin B^{2\varepsilon}$, then

$$\lambda(\underset{\sim}{x} - \underset{\sim}{y};\, B) = 0 \quad \text{whenever} \quad |\underset{\sim}{y}| \leq 2\varepsilon .$$

Hence $M(\underset{\sim}{x};\, B) = 0$ for $\underset{\sim}{x} \notin B^{2\varepsilon}$ i.e.

33

$$(5.7.15) \qquad M(\underset{\sim}{x}\,;\, B) \leq \Lambda(\underset{\sim}{x}\,;\, B^{2\varepsilon})$$

The lemma is then obtained from (5.7.14) and (5.7.15) by putting $B = A^{-2\varepsilon}$ and $B = A^{2\varepsilon}$. The proof is thus complete.

Lemma 5.7.2. Let μ_1 be a probability distribution and μ_2 a signed measure on R_k. Let $p_1(\underset{\sim}{t})$ and $p_2(\underset{\sim}{t})$ denote their Fourier transforms. Suppose A is a convex set such that (1) the inradius $r(A) > 4\varepsilon$ and (2) $A \subset S(o, m)$. Then

$$(5.7.16) \qquad |\mu_1(A) - \mu_2(A)| \leq \frac{4\pi^{(k-1)/2}\,(m + 2\varepsilon)^{k-1}}{\Gamma(\frac{1}{2} + \frac{1}{2}k)}\, I +$$
$$+ \operatorname{Max}\left\{ |\mu_2|(A^{4\varepsilon} - A),\ |\mu_2|(A - A^{-4\varepsilon}) \right\}$$

where

$$(5.7.17) \qquad I = \frac{1}{(2\pi)^k} \int |q_\varepsilon(\underset{\sim}{t})|^2 \, \frac{|p_1(\underset{\sim}{t}) - p_2(\underset{\sim}{t})|\, d\underset{\sim}{t}}{|\underset{\sim}{t}|}$$

and $|\mu_2|$ denotes the absolute variation of μ_2.

Proof. From the inequality (5.7.12) it follows that

$$(5.7.18) \qquad \mu_1(A) - \mu_2(A) \leq \int M(\underset{\sim}{x};\, A^{2\varepsilon})\, d\mu_1 - \mu_2(A)$$
$$\leq \int M(\underset{\sim}{x};\, A^{2\varepsilon})\, d(\mu_1 - \mu_2) +$$
$$\int M(\underset{\sim}{x},\, A^{2\varepsilon})\, d\mu_2 - \mu_2(A).$$

34 If we utilize (5.7.12) once again it follows that

$$(5.7.19)\quad \int M(\underset{\sim}{x}, A^{2\varepsilon})d\mu_2 - \mu_2(A) \le \int [M(\underset{\sim}{x}, A^{2\varepsilon}) - \chi(\underset{\sim}{x}, A)]d|\mu_2|$$

$$\le \int [\chi(\underset{\sim}{x}, A^{4\varepsilon}) - \chi(\underset{\sim}{x}, A)]d|\mu_2|$$

$$= |\mu_2|(A^{4\varepsilon} - A).$$

Thus we have from (5.7.18) and (5.7.19) that

$$(5.7.20)\quad \mu_1(A) - \mu_2(A) \le \int M(\underset{\sim}{x}, A^{2\varepsilon})\, d(\mu_1 - \mu_2) + |\mu_2|(A^{4\varepsilon} - A).$$

From the inequality (5.7.11) it follows that

$$\mu_1(A) - \mu_2(A) \ge \int M(\underset{\sim}{x}, A^{-2\varepsilon})d\mu_1 - \mu_2(A)$$

$$\ge \int M(\underset{\sim}{x}, A^{-2\varepsilon})\, d(\mu_1 - \mu_2) +$$

$$\int [M(\underset{\sim}{x}, A^{-2\varepsilon}) - \chi(\underset{\sim}{x}, A)]d\mu_2$$

$$\ge \int M(\underset{\sim}{x}, A^{-2\varepsilon})d(\mu_1 - \mu_2)$$

$$- \int [\chi(\underset{\sim}{x}, A) - M(\underset{\sim}{x}, A^{-2\varepsilon})]d|\mu_2|$$

since by 5.7.11, $\chi(x; A) \ge M(x, A^{-2\varepsilon})$. Once again utilizing (5.7.11) we obtain

$$(5.7.21)\quad \mu_1(A) - \mu_2(A) \ge \int M(\underset{\sim}{x}, A^{-2\varepsilon})d(\mu_1 - \mu_2) - |\mu_2|(A - A^{-4\varepsilon}).$$

Combining (5.7.20) and (5.7.21) we have

$$(5.7.22)\quad |\mu_1(A) - \mu_2(A)| \leq \text{Max}\,(|I_1|, |I_2|) + \text{Max}\left\{|\mu_2|(A^{4\varepsilon} - A),\ |\mu_2|(A - A^{-4\varepsilon})\right\}$$

where

$$I_1 = \int M(\underset{\sim}{x},\ A^{2\varepsilon})\ d(\mu_1 - \mu_2)$$

and

$$I_2 = \int M(\underset{\sim}{x},\ A^{-2\varepsilon})\ d(\mu_1 - \mu_2)$$

Since the Fourier transform $m(\underset{\sim}{t},\ A^{2\varepsilon})$, is Lebesgue integrable (cf. 5.7.10) it follows from the Fourier inversion theorem that

$$(5.7.23)\quad I_1 = \frac{1}{(2\pi)^k} \int m(\underset{\sim}{t},\ A^{2\varepsilon})[p_1(\underset{\sim}{t}) - p_2(\underset{\sim}{t})]\ d\underset{\sim}{t}$$

$p_1(\underset{\sim}{t})$, $p_2(\underset{\sim}{t})$ being the Fourier transforms of μ_1 and μ_2 respectively. Also, since $A^{2\varepsilon} \subset S(0,\ m+2\varepsilon)$ it follows from Theorem 4.6.1. of Appendix A that

$$(5.7.24)\quad |m(\underset{\sim}{t},\ A^{2\varepsilon})| \leq |q(\underset{\sim}{t},\ A^{2\varepsilon})| \leq \frac{4}{|\underset{\sim}{t}|}\ \frac{\pi^{\frac{1}{2}(k-1)}\ (m + 2\varepsilon)^{k-1}}{\Gamma\left(\frac{1}{2}\ (k+1)\right)}.$$

From (5.7.23) and (5.7.24) it follows that

36 (5.7.25) $$|I_1| \leq 4\,\pi^{\frac{1}{2}(k-1)}\ \frac{(m+2\varepsilon)^{k-1}}{\Gamma(\frac{1}{2}(k+1))}\ I$$

where I is defined by (5.7.17). Similarly

(5.7.26) $$|I_2| \leq 4\,\pi^{\frac{1}{2}(k-1)}\ \frac{(m-2\varepsilon)^{k-1}}{\Gamma(\frac{1}{2}(k+1))}\ I$$

The lemma is an immediate consequence of (5.7.22), (5.7.25) and (5.7.26).

The proof of the lemma is complete.

5.8. Proof of Theorem 5.3.1. Let m and ε be positive quantities whose values will be specified later. Let

$$\Delta_n(A) = |P_n(A) - \Phi(A)|, \text{ and}$$

$$\Delta_n = \sup_{A \in \mathcal{C}} |\Delta_n(A)|$$

(all sets considered in the rest of the Chapter are convex sets). The quantities $C_1, C_2, \ldots$ denote constants depending only on k.

The proof will be carried out in three steps. The first step consists in reducing the problem to sets contained in $S(o,m)$. The following inequality is easily established

(5.8.1) $$\Delta_n \leq \sup_{A \subset S(o,m)} \Delta_n(A) + (P_n + \Phi)(R_k - S(o, m)).$$

It is clear that

$$[x : |x_i| > m/\sqrt{k}\ ;\ i = 1, 2, \ldots, k] \supset S(o, m)$$

Consequently

$$(5.8.2) \qquad P_n(R_k - S(o, m)) \leq \sum_{j=1}^{k} P_n\ (|x_j| > m/\sqrt{k})$$

It follows from the Berry-Esseen inequality (Esseen [15], p. 43) that

$$(5.8.3) \qquad P_n\ (|x_i| > m/\sqrt{k}) = P\ [\ |Y_j^{(n)}| > m/\sqrt{k}\ ]$$

$$\leq C_1\ \frac{E|Z_j^{(1)}|^3}{\sqrt{n}} + \int_{|x| > m/\sqrt{k}} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}x^2}\ dx$$

$$\leq C_1\ \frac{\beta_{3j}}{\sqrt{n}} + \frac{\sqrt{k}}{\sqrt{2\pi}\, m}\ \exp\ (-\frac{1}{2k}\ m^2)$$

It is easily verified that

$$(5.8.4) \qquad \frac{1}{(\sqrt{2\pi})^k} \int_{|x| > m} e^{-\frac{1}{2}|x|^2}\ dx \leq C_2(\frac{1}{m})\ \exp\ (-\frac{1}{2k}\ m^2)$$

Combining (5.8.1) - (5.8.4) we obtain

38

$$(5.8.5) \qquad \Delta_n \le \sup_{A \subset S(o,m)} \Delta_n(A) + \frac{C_1 \beta_3}{\sqrt{n}} + C_3 \frac{e^{m^2/4}}{m} \exp\left(-\frac{m^2}{2k}\right)$$

Now suppose $A \subset S(o, m)$. Then there are two possibilities. Either $r(A) \le 4\varepsilon$ or $r(A) > 4\varepsilon$, where $r(A)$ denotes the inradius of A. Let us first consider sets A for which $r(A) \le 4\varepsilon$. In this case it follows from Theorem A.2.1 (cf. Appendix A) that there exist a direction vector $\underset{\sim}{l} = (l_1, \ldots, l_k)$ $(\sum l_i^2 = 1)$ such that

$$A \subset [\underset{\sim}{x} ; \; a \le \underset{\sim}{l}' . \underset{\sim}{x} \le b]$$

where

$$| a - b | < 2 \sqrt{k+2} \; r(A) \le 8 \sqrt{k+2} \; \varepsilon .$$

If we apply the Berry-Esseen inequality once again to the normed sum $\underset{\sim}{l}' . Y^{(n)} = (\underset{\sim}{l}' . Z^{(1)} + \ldots + \underset{\sim}{l}' . Z^{(n)}) / \sqrt{n}$, it follows that

$$| P_n(A) - \Phi(A) | \le 2C_1 \cdot \frac{E| Z^{(1)}|^3}{\sqrt{n}} + \frac{16 \sqrt{k+2} \; \varepsilon}{\sqrt{2\pi}}$$

Since $| \underset{\sim}{l}' Z^{(1)} | \le | Z^{(1)} |$, this reduces to the inequality

$$(5.8.6) \qquad \sup_{r(A) \le 4\varepsilon} \Delta_n(A) \le \frac{2C_1 \beta_3}{\sqrt{n}} + C_4 \varepsilon .$$

Now consider sets $A \subset S(o, m)$ for which $r(A) > 4\varepsilon$. If we put $\mu_1 = P_n$ and $\mu_2 = \Phi + n^{-\frac{1}{2}} \tilde{P}_1 (-\Phi)$, in Lemma 5.7.2, we have

$$(5.8.7)\quad \Delta_n(A) \leq C_5\, m^{k-1}\, I + \operatorname{Max}\left\{ |\mu_2|(A^{4\varepsilon} - A), |\mu_2|(A - A^{-4\varepsilon}) \right\}$$

$$+ \frac{1}{\sqrt{n}} \left| \int_A d\, \tilde{P}_1(-\Phi) \right|$$

where

$$(5.8.8)\quad I = \int \frac{|q_\varepsilon(t)|^2}{|t|} \left| f_n(t) - e^{-\frac{1}{2}|t|^2} (1 + n^{-\frac{1}{2}} \tilde{P}_1(it)) \right| dt .$$

It is then easy to deduce from the definition (5.5.1) of $\tilde{P}_1(-\Phi)$ that

$$(5.8.9)\quad |\mu_2|(A^{4\varepsilon} - A) \leq 3\vartheta_3 \int_{A^{4\varepsilon}-A} (1 + |x| + |x|^3) e^{-\frac{1}{2}|x|^2} dx$$

$$\leq C_6\, \vartheta_3\, \varepsilon$$

In (5.8.9) the last step is a consequence of lemma A.5.2 of appendix A. Similarly,

$$(5.8.10)\qquad |\mu_2|(A - A^{-4\varepsilon}) \leq C_7\, \vartheta_3\, \varepsilon .$$

Once again it is easy to deduce from (5.5.2) that

40

$$(5.8.11)\qquad \Big| \int_A d\tilde{P}_1(-\underset{\sim}{0}) \Big| \le 39_3 \int (|\underset{\sim}{x}| + |\underset{\sim}{x}|^3) e^{-\frac{1}{2}|\underset{\sim}{x}|^2}\, d\underset{\sim}{x} \le C_8 9_3 .$$

It thus remains only to estimate the integral I defined by (5.8.8). In view of (5.7.5) and (5.7.8) we have

$$(5.8.12)\qquad I \le \int_{|\underset{\sim}{t}| < T_{sn}} |\underset{\sim}{t}|^{-1} \Big| f_n(\underset{\sim}{t}) - e^{-\frac{1}{2}|\underset{\sim}{t}|^2} (1+n^{-\frac{1}{2}}\tilde{P}_1(i\underset{\sim}{t})) \Big| d\underset{\sim}{t} +$$

$$+ \frac{C_9}{\varepsilon^{(k+1)}} \int_{|\underset{\sim}{t}| > T_{sn}} \frac{|f_n(\underset{\sim}{t})|}{|\underset{\sim}{t}|^{k+2}}\, d\underset{\sim}{t} + \int_{|\underset{\sim}{t}| > T_{sn}} e^{-\frac{1}{2}|\underset{\sim}{t}|^2} \Big| (1+n^{-\frac{1}{2}}\tilde{P}_1(i\underset{\sim}{t})) \Big| d\underset{\sim}{t}$$

$$= I_1 + I_2 + I_3 \text{ say,}$$

where $T_{sn} = \sqrt{n}\ 9_4^{-3/4}/32$. It follows from (5.4.7) that

$$(5.8.13)\qquad I_1 \le \frac{C_{10}\, 9_4^{3/2}}{n} \le \frac{C_{11}\, \beta_4^{3/2}}{n} .$$

It is easy to see that

$$(5.8.14)\qquad I_3 \le \frac{C_{12}\, \beta_4^{3/2}}{n^{3/2}}$$

(actually I_3 is of exponential order but the above inequality suffices for our purpose). If we change the variable of integration in I_2 to $\underset{\sim}{t}\sqrt{n}$,

We obtain

$$(5.8.15)\qquad I_2 \le \frac{C_9}{\varepsilon^{k+1}}\,\frac{1}{n}\int_{|\underset{\sim}{t}|>(32\,\varrho_4^{3/4})^{-1}} \frac{|f(\underset{\sim}{t})|^n}{|\underset{\sim}{t}|^{k+2}}\,d\underset{\sim}{t}.$$

Theorem 5.6.2. now implies that

$$(5.8.16)\qquad \int_{|\underset{\sim}{t}|>(32\,\varrho_4^{3/4})^{-1}} |\underset{\sim}{t}|^{-(k+2)}\,|f(\underset{\sim}{t})|^n d\underset{\sim}{t} \le \frac{C_{13}}{n^{\frac{1}{2}k}}\,\beta_4^{9/4}.$$

Combining (5.8.15) and (5.8.16) we have

$$(5.8.17)\qquad I_2 \le C_{14}\,\frac{\beta_4^{9/4}}{\varepsilon^{k+1}\,n^{(k+2)/2}}.$$

From (5.8.12) - (5.8.14) and (5.8.17) we deduce that

$$(5.8.18)\qquad I \le C_{15}\left\{\frac{\beta_4^{3/2}}{n} + \frac{\beta_4^{3/2}}{n^{3/2}} + \frac{\beta_4^{9/4}}{\varepsilon^{k+1}\,n^{(k+2)/2}}\right\}.$$

Pooling together the inequalities (5.8.7),(5.8.9)-(5.8.11) and (5.8.18) we obtain

$$(5.8.19)\qquad \sup_{A\subset S(o,m)} \Delta_n(A) \le C_{16}\left[\varrho_3\varepsilon + \frac{\varrho_3}{\sqrt{n}} + m^{k-1}\left\{\frac{\beta_4^{3/2}}{n} + \frac{\beta_4^{3/2}}{n^{3/2}} + \frac{\beta_4^{9/4}}{\varepsilon^{k+1}\,n^{(k+2)/2}}\right\}\right].$$

42 The final estimate of Δ_n is obtained from (5.8.5), (5.8.6) and (5.8.19). We now put

$$\varepsilon = \frac{\beta_4^{3/4} (k \log n)^{\alpha}}{\sqrt{n}} , \quad m = \sqrt{k \log n} \tag{5.8.20}$$

where

$$\alpha = \frac{(k-1)}{2(k+1)} \tag{5.8.21}$$

Then by elementary calculations it can be deduced from (5.8.5), (5.8.6), (5.8.19) and (5.8.20) that

$$\Delta_n \leq C_{17} \frac{\beta_4^{3/2} (\log n)^{\alpha}}{\sqrt{n}}$$

where α is determined by (5.8.21). This completes the proof of Theorem 5.3.1.

5.9. Proof of Theorem 5.3.2. The proof of this theorem is analogous to that of 5.3.1. We present therefore only a brief sketch of the proof.

Let

$$\Psi_n^{(s)}(x) = \Phi(\underset{\sim}{x}) + \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} \tilde{P}_j(-\Phi) \tag{5.9.1}$$

where the functions $\tilde{P}_j(-\infty)$ are defined by (5.5.1). Let

$$\Delta_n^{(s)}(A) = |P_n(A) - \int_A d\,\Psi_n^{(s)}(\underset{\sim}{x})|$$

and

$$\Delta_n^{(s)} = \sup_{A \in \mathcal{C}} \Delta_n^{(s)}(A)$$

where $\mathcal{C}$ denotes the class of all convex sets. Let ε, m be positive real numbers to be specified later. Let $M_1, M_2, \ldots$ denote constants depending only on distribution of $Z^{(1)}$.

To begin with, let us note that the c.f. of the real valued random variable ξ

$$\xi = \underset{\sim}{l}' \cdot Y^{(n)} = \sum_{j=1}^{k} l_j\, Y_j^{(n)},$$

where $\underset{\sim}{l}$ is a direction vector, also satisfies the condition (C). Consequently it follows from known results (cf. Cramér [10], p. 84; Esseen [15], p. 48) that the distribution function $G_n(x)$ of ξ admits the following expansion:

$$(5.9.2)\quad G_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2}\,dx + \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\left[\frac{K_1(x)}{\sqrt{n}} + \ldots + \frac{K_{s-2}(x)}{n^{\frac{1}{2}(s-2)}}\right] + o\left(n^{-\frac{1}{2}(s-2)}\right)$$

the $K_j(x)$ for $j = 1, 2, \ldots (s-2)$ are polynomials in x involving only the first s cumulants of ξ. If we utilize the above expansion instead of the Berry-Esseen inequality we can obtain in-equalities similar to (5.8.5) and (5.8.6). In fact following exactly the same method as in the last section it is not difficult to establish the following two inequalities:

$$(5.9.3) \qquad \Delta_n^{(s)} \leq \sup_{A \subset S(o,m)} \Delta_n^{(s)}(A) + M_1 e^{-\frac{m^2}{4k}} + O(n^{-\frac{1}{2}(s-2)})$$

$$(5.9.4) \qquad \sup_{r(A) \leq 4\varepsilon} \Delta_n^{(s)}(A) \leq M_2 \varepsilon + O(n^{-\frac{1}{2}(s-2)}).$$

It thus remains to estimate $\Delta_n^{(s)}(A)$ for sets A which are such that $A \subset S(o, m)$ and $r(A) > 4\varepsilon$. It follows from Lemma 5.7.2 with

$$\mu_1 = P_n \text{ , and } \mu_2(A) = \int_A d\,\Psi_n^{(s)}$$

where $\Psi_n^{(s)}$ is defined by (5.9.1), that

$$(5.9.4) \qquad \Delta_n^{(s)}(A) \leq C_1 m^{k-1} I + \text{Max} \left\{ |\mu_2|(A^{4\varepsilon} - A), |\mu_2|(A - A^{-4\varepsilon}) \right\}$$

where

45

$$(5.9.5) \qquad I = \int |q_\varepsilon(\underset{\sim}{t})|^2 \, \frac{|f_n(\underset{\sim}{t}) - \psi_n^{(s)}(\underset{\sim}{t})|\, d\underset{\sim}{t}}{|\underset{\sim}{t}|}$$

and $|\mu_2|$ denotes the absolute variation of μ_2. Here $\psi_n^{(s)}(\underset{\sim}{t})$ is the Fourier transform of μ_2. From (5.9.1) and Lemma 5.5.1, we have

$$(5.9.6) \qquad \psi_n^{(s)}(\underset{\sim}{t}) = \sum_{j=0}^{s-3} n^{-\frac{1}{2}j} \tilde{P}_j(i\underset{\sim}{t})\, e^{-\frac{1}{2}|\underset{\sim}{t}|^2}.$$

It follows easily from the definition (5.9.1) of $\Psi_n^{(s)}$, that

$$|\mu_2|(A) \le M_3 \int_A e^{-\frac{1}{4}|\underset{\sim}{x}|^2} \, d\underset{\sim}{x}.$$

Hence, by Lemma A.5.2 of Appendix A,

$$(5.9.7) \qquad |\mu_2|\,(A^{4\varepsilon} - A) \le M_4 \varepsilon$$

$$|\mu_2|\,(A - A^{-4\varepsilon}) \le M_5 \varepsilon$$

To estimate I, given by (5.9.5), let

$$(5.9.8) \qquad T_{sn} = \sqrt{n}\,(8s\,\varrho_s^{3/s})^{-1}.$$

Then we can write

46

$$I = \int_{|\underset{\sim}{t}| < T_{sn}} |q_\varepsilon(\underset{\sim}{t})|^2 \frac{|f_n(\underset{\sim}{t}) - \psi_n^{(s)}(\underset{\sim}{t})|}{|\underset{\sim}{t}|} d\underset{\sim}{t} +$$

$$+ \int_{|\underset{\sim}{t}| > T_{sn}} |q_\varepsilon(\underset{\sim}{t})|^2 |\underset{\sim}{t}|^{-1} |f_n(\underset{\sim}{t})| d\underset{\sim}{t}$$

$$+ \int_{|\underset{\sim}{t}| > T_{sn}} |q_\varepsilon(\underset{\sim}{t})|^2 |\underset{\sim}{t}|^{-1} |\psi_n^{(s)}(\underset{\sim}{t})| d\underset{\sim}{t}$$

$$= I_1 + I_2 + I_3 \quad \text{say.}$$

Then it follows from (5.4.7) and (5.7.5) that

$$(5.9.9) \qquad I_1 = O(n^{-\frac{1}{2}(s-2)}).$$

If we change the variable of integration in I_2 from $\underset{\sim}{t}$ to $\underset{\sim}{t}\sqrt{n}$, it follows from (5.7.8) that

$$(5.9.10) \qquad I_2 \leq \frac{M_6}{n\varepsilon^{k+1}} \int_{|\underset{\sim}{t}| > (8s\varrho_s^{3/s})^{-1}} |f(\underset{\sim}{t})|^n |\underset{\sim}{t}|^{-(k+2)} d\underset{\sim}{t}.$$

Let

$$\delta = \sup \left[|f(\underset{\sim}{t})| ; \ |\underset{\sim}{t}| > (8s\varrho_s^{3/s})^{-1} \right].$$

Since the condition (C) (5.3.2) is satisfied we have

$$0 < \delta < 1.$$

Hence it follows from (5.9.10) that

$$(5.9.11) \qquad I_2 \leq M_7 \delta^n \frac{1}{n} \frac{1}{\varepsilon^{k+1}} .$$

It is easy to show that

$$(5.9.12) \qquad I_3 \leq M_8 \int_{|\underset{\sim}{t}| > T_{sn}} e^{-\frac{1}{4}|\underset{\sim}{t}|^2} d\underset{\sim}{t} \leq M_9 \frac{1}{\sqrt{n}} e^{-\frac{1}{4}T_{sn}^2}$$

Combining (5.9.9), (5.9.11) and (5.9.12) we obtain

$$(5.9.13) \qquad I \leq M_7 m^{k-1} \left[\frac{1}{n\varepsilon^{k+1}} \delta^n + O(n^{-\frac{1}{2}(s-2)}) \right].$$

Pooling together the inequalities (5.9.4), (5.9.7) and (5.9.13) we have

$$(5.9.14) \qquad \sup_{A \subset S(o,m)} \Delta_n(A) \leq M_8\varepsilon + M_7 m^{k-1} \left[\frac{\delta^n}{n\varepsilon^{k+1}} + O(n^{-\frac{1}{2}(s-2)}) \right].$$

Thus it follows from (5.9.2), (5.9.3) and (5.9.14) that

$$(5.9.15) \qquad \Delta_n \leq M_9\varepsilon + \frac{M_1}{m} e^{-m^2/4k} + M_7 m^{k-1} \left[\frac{\delta^n}{n\varepsilon^{k+1}} + O(n^{-\frac{1}{2}(s-2)}) \right].$$

48

If we put

$$\varepsilon = n^{-\frac{1}{2}(s-2)}, \text{ and } \quad m = (2k(s-2)\ \log n)^{\frac{1}{2}},$$

in (5.9.15), we have

$$\Delta_n = O(n^{-\frac{1}{2}(s-2)}\ (\log n)^{(k-1)/2})$$

This completes the proof of theorem 5.3.2.

CHAPTER 6

On the Central Limit Theorem in R_k. Asymptotic expansions in the case of Lattice Distributions.

Introduction. Let $Z^{(1)}, Z^{(2)}, \ldots$ be a sequence of independent and identically distributed random vectors with mean zero and variance covariance matrix I. If P_n is the probability distribution of their normed sum

$$Y^{(n)} = \frac{Z^{(1)} + \ldots + Z^{(n)}}{\sqrt{n}},$$

then it was shown in the last Chapter (Theorem 5.3.2) that the following expansion is valid:

$$(6.1.1) \qquad P(A) = \int_A d\,\Phi + \frac{1}{\sqrt{n}} \int_A d\,\tilde{P}_1(-\Phi) + o\left(\frac{1}{n}\right)$$

for all convex sets A. It should be noted that $\tilde{P}_1(-\Phi)$ (defined by 5.5.2) is absolutely continuous with respect to Lebesgue measure in R_k. For the validity of (6.1.1) it is necessary to suppose, apart from the existence of moments, that the following condition (Cramér's condition) is satisfied:

$$(6.1.2) \qquad \limsup_{|\underset{\sim}{t}| \to \infty} |f(\underset{\sim}{t})| < 1$$

where $f(\underset{\sim}{t})$ is the c.f. of $Z^{(1)}$. Since there are a large class of distributions which do not fulfill (6.1.2), for instance the class of all discrete distributions, an asymptotic expansion of the error valid for these distributions as well, becomes important. That is the main object of investigation in this chapter.

In the first place it should be noted that the expansion (6.1.1) is not in general valid if (6.1.2) is not satisfied. This becomes clear from the fact that while $F_n(x)$ has possibly jumps of the order of magnitude of $n^{-\frac{1}{2}}$, the r.h.s has no jumps at all. The same situation also persists if $k = 1$. In this case Esseen ([15]) obtained an expansion for a large class of distributions known as lattice distributions. For these he showed that

$$(6.1.2) \qquad F_n(x) = \Phi(x) + \frac{1}{\sqrt{n}} \tilde{P}_1(-\Phi) + \frac{h}{\sqrt{n}} S_1(x)\, e^{-\frac{1}{2}x^2} + o\left(\frac{1}{n}\right)$$

where $S_1(x)$ is a certain discontinuous functions. The lattice distributions are those for which $|f(t)|$ is a periodic function of t and consequently violate the condition (6.1.2) in the worst possible way. Analogously we consider lattice distributions in R_k, in section 2 of this chapter. For these distributions we obtain an asymptotic expansion, if $\beta_5 < \infty$

$$(6.1.3) \qquad P_n(A) = \int_A dQ(x) + \frac{1}{\sqrt{n}} \int_A d\tilde{P}_1(-\Phi) + o\left(\frac{1}{n}\right)$$

where

$$Q(\underset{\sim}{x}) = \frac{k}{\prod_{j=1}} [1 - n^{-\frac{1}{2}} \; S_1[\, x_j \; \sqrt{n} - na_j] \; \frac{\partial}{\partial x_j}] \; \Phi(x)$$

and the order of the remainder term in (6.1.3) is uniformly $O(n^{-1})$ over all Borel sets (Theorem 6.4.4). A feature of the above result is that the expansion (6.1.3) is valid for all Borel sets, a result which does not seem to have been observed even in the case $k = 1$. Thus if A is an infinite rectangle, say $A = [(-\infty, x_i), i = 1,2,..,k]$ then

$$(6.1.4) \qquad F_n(x_1, \ldots, x_k) = \Phi(x) + \frac{1}{\sqrt{n}} \; \tilde{P}_1 \; (-\Phi) +$$

$$\frac{1}{\sqrt{n}} \; \sum_{j=1}^{k} S_1[x_j \; \sqrt{n} - na_j] \; \frac{\partial \Phi}{\partial x_j} + O\left(\frac{1}{n}\right)$$

where $F_n(\underset{\sim}{x})$ is the d.f. of the random vector $Z^{(1)}$ (Theorem 6.4.2).

The method of this chapter consists in obtaining the local expansion first, and then deducing the integrated expansion. This process is actually carried out by the application of the generalization of Euler-Maclaurin sum formula to k dimensions given in Appendix B.

4 6.2. Lattice Distributions in R_k. In this section the lattice distributions are defined and their properties are discussed. All the material presented here is well known and may be found in Esseen (cf. [15]). However the exposition is given here in some detail for the sake of continuity, ease of reference, and to bring out certain points not emphasized by Esseen.

In what follows elements of R_k will denoted by $\underset{\sim}{x}$, $\underset{\sim}{y}$, etc. By an integer vector $\underset{\sim}{m}$ we mean a vector $\underset{\sim}{m} = (m_1, \ldots, m_k)$ where each m_i is an integer ; positive or negative or zero . A set of points of the form

$$\{\underset{\sim}{z}_o + H\underset{\sim}{m}\} \tag{6.2.1}$$

where $\underset{\sim}{m}$ ranges over all integer vectors and H is a non-singular matrix, is called a lattice. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are two lattices, then $\mathcal{L}_1$ is said to be a sub-lattice of $\mathcal{L}_2$ if $\mathcal{L}_1 \subset \mathcal{L}_2$.

Definition 6.2.1. Let Z be a random vector and F its distribution function. Then F is said to be a lattice distribution if there is a lattice $\mathcal{L}$ such that Z is confined to $\mathcal{L}$ with probability one.

Let $f(\underset{\sim}{t})$ be the c.f. of the lattice vector Z. Then

$$f(\underset{\sim}{t}) = \sum_{\underset{\sim}{m}} p(\underset{\sim}{m})\, e^{i\, \underset{\sim}{t}'.(\underset{\sim}{z}_o + H\underset{\sim}{m})} \tag{6.2.2}$$

where the summation extends over all integer vectors and

$$p(\underset{\sim}{m}) = P\left[Z = \underset{\sim}{z}_0 + H\underset{\sim}{m} \right]$$

From (6.2.2) it at once follows that

$$(6.2.3) \qquad | f(\underset{\sim}{t} + 2\pi H^{-1} \underset{\sim}{\nu}) | = | f(\underset{\sim}{t}) |$$

where $\underset{\sim}{\nu}$ is any integer vector. In other words $| f(\underset{\sim}{t}) |$ is a periodic function with k-linearly independent periods. Conversely it is not difficult to verify that if $| f(\underset{\sim}{t}) |$ is periodic with k- linearly independent periods then Z is a lattice vector (has lattice d.f.)

A question of some importance in our analysis is the value distribution of $| f(\underset{\sim}{t}) |$, in particular the points at which $| f(\underset{\sim}{t}) | = 1$. It is clear from (6.2.3) that this set includes the lattice of points

$$(6.2.4) \qquad \left\{ 2\pi H^{-1}\underset{\sim}{\nu} ; \ \underset{\sim}{\nu} \text{ is an integer vector} \right\}.$$

However there may be points other than elements of the lattice (6.2.4) at which $| f(\underset{\sim}{t}) | = 1$. The existence of these points is accounted by the fact that Z may actually be concentrated in a sub-lattice of (6.2.1), such as for instance, the set of points

$$\left\{ \underset{\sim}{z}_0 + 2\, H\underset{\sim}{m} \right\}$$

where $\underset{\sim}{m}$ is an integer vector. There is thus in general a certain

6 non-uniquess about the lattice (6.2.1) in which Z is concentrated. This can however be rectified by the introduction of a concept analogous to that of the maximum span in the one-dimensional case.

For any lattice random vector Z we say that a lattice $\mathcal{L}$ is minimal if (1) Z is confined to $\mathcal{L}$ with probability one and (2) there is no proper sub-lattice $\mathcal{L}_1 \subset \mathcal{L}$ such that $P[Z \in \mathcal{L}_1] = 1$.

The existence of such a minimal lattice may be shown as follows. Let

$$\mathcal{L} = \{ \underset{\sim}{z}_0 + H \underset{\sim}{m} \}$$

be an arbitrary lattice in which Z is concentrated. Let

$$Y = H^{-1} (Z - \underset{\sim}{z}_0)$$

It is then clear that there is a one-to-one correspondence between the lattices in which Y and Z are concentrated. Also this correspondence preserves minimality. It thus suffices to consider the random vector Y. Clearly from its definition Y is itself is concentrated in the lattice of all integral vectors. Consequently the projections $Y_1 \ldots Y_k$ on the axes are themselves real valued lattice variables. Let $d_1 \ldots d_k$ be their corresponding maximum spans and D a diagonal matrix with

7

$d_1 \ldots d_k$ as diagonal elements. Then Y is concentrated in the lattice

$$\{ D \underset{\sim}{m} \}$$

which also happens to be minimal as is easily verified.

We shall suppose henceforth that the lattice (6.2.1) associated with Z is minimal. Then $| f (\underset{\sim}{t}) | = 1$ if and only if

$$\underset{\sim}{t} \in \{ 2 \pi H^{-1} \underset{\sim}{m} \}$$

The proof is easily carried out by considering the random vector Y used in the last paragraph. We shall omit the details.

Let $\mathcal{H}$ denote the (k-dimensional) parallogram formed by the row vectors of $2\pi H^{-1}$. This is called, for obvious reasons, the parallogram of periodicity. The minimal character of the lattice (6.2.1) now implies that if $\mathcal{H}_0$ denotes the parallogram $\mathcal{H}$ moved in such a way that its centre and the origin coincide, then at all points $\underset{\sim}{t} \neq 0$ in the interior or on the boundary of $\mathcal{H}_0$ $| f (\underset{\sim}{t}) | < 1$. Consequently given any $\delta > 0$, there exists a $\varrho = \varrho(\delta)$ such that for all $\underset{\sim}{t} \in \mathcal{H}_0$ and $| \underset{\sim}{t} | \geq \delta$

$$| f (\underset{\sim}{t}) | \leq \varrho < 1. \tag{6.2.5}$$

We conclude this section with the statement of a well known formula connecting $p(\underset{\sim}{m})$ and the cf. $f(\underset{\sim}{t})$. Multiply both sides of (6.2.2) by

$$\exp\left[-i\,\underset{\sim}{t}'.(\underset{\sim}{z}_0 + H\underset{\sim}{m})\right]$$

and integrate over the region $\mathcal{H}_0$. Since

$$\int_{\mathcal{H}_0} e^{i\,\underset{\sim}{t}'.H(\underset{\sim}{m}-\underset{\sim}{n})}\,d\underset{\sim}{t} = \frac{1}{|H|}\int_{[|\tau_r|\le\pi\,;r=1,\ldots k]} e^{i\,\underset{\sim}{\tau}'.(\underset{\sim}{m}-\underset{\sim}{n})}\,d\underset{\sim}{\tau}$$

$$= \begin{cases} \dfrac{(2\pi)^k}{|H|} & \text{if } \underset{\sim}{m} = \underset{\sim}{n} \\ 0 & \text{if } \underset{\sim}{m} \neq \underset{\sim}{n}\,, \end{cases}$$

we obtain as a consequence

$$(6.2.6)\qquad p(\underset{\sim}{m}) = \frac{|H|}{(2\pi)^k}\int_{\mathcal{H}_0} f(\underset{\sim}{t})\exp\left[-i\,\underset{\sim}{t}'.(H\underset{\sim}{m}+\underset{\sim}{z}_0)\right]d\underset{\sim}{t}$$

6.3. Local limit theorems and estimation of the remainder. Let $Z^{(1)}, Z^{(2)}, \ldots$ be a sequence of independent random vectors in R_k with a common lattice distribution. We suppose that $Z^{(1)}$ is confined to the set of lattice points

$$(6.3.1)\qquad \left\{ \underset{\sim}{a} + \underset{\sim}{m}\ ;\ \underset{\sim}{m} \text{ an integer vector} \right\}$$

with probability one. Since any lattice can be reduced to the form

(6.3.1) by a linear transformation, there is no loss of generality involved in our assumption. We also suppose that

(i) the mean value of each component is zero, i.e. $E\, Z_j^{(1)} = 0$ for $j = 1, 2, \ldots k$.

(ii) the lattice (6.3.1) is the minimal lattice in which $Z^{(1)}$ is concentrated.

In addition to the above we will suppose the existence of higher order moments; in any case, moments of the second order. Let V denote the variance-covariance matrix of $Z^{(1)}$.

Let

$$Y^{(n)} = [\, Z^{(1)} + \ldots + Z^{(n)}] / \sqrt{n} \,.$$

Clearly $Y^{(n)}$ is also a lattice vector and its possible values are restricted to the lattice

$$(6.3.2) \qquad \left\{ (n\underset{\sim}{a} + \underset{\sim}{m}) / \sqrt{n} \right\}$$

where $\underset{\sim}{m}$ is an arbitrary integer vector. Let $p_n(\underset{\sim}{z})$ denote the probability mass at a discontinuity point $\underset{\sim}{z}$ belonging to the lattice (6.2.2) i.e.

$$p_n(\underset{\sim}{z}) = P [\, Y^{(n)} = \underset{\sim}{z} \,]$$

In this section we study the probabilities $p_n(\underset{\sim}{z})$ and obtain an asymptotic expansion for $p_n(\underset{\sim}{z})$. This type of problem concerning the behaviour of $p_n(\underset{\sim}{z})$ is sometimes called a local limit problem.

We begin with a well known result. This result gives an asymptotic expansion for $p_n(\underset{\sim}{z})$ in terms of $\varphi(\underset{\sim}{x})$ and its derivatives (cf. Esseen [15], p 116). Since its proof is entirely analogous to the corresponding result for the case $k = 1$ we omit the proof.

Theorem 6.3.1. Let $\beta_s < \infty$. Then for every $\underset{\sim}{z}$ belonging to the lattice (6.3.2) and uniformly in $\underset{\sim}{z}$ we have

$$(6.3.3) \qquad p_n(\underset{\sim}{z}) = q_n^{(s)}(\underset{\sim}{z}) + o\left(n^{-\frac{1}{2}(s+k-2)}\right)$$

where

$$(6.3.4) \qquad q_n^{(s)}(\underset{\sim}{x}) = n^{-\frac{1}{2}k} \sum_{j=1}^{s-3} n^{-\frac{1}{2}j} P_j(-\varphi)$$

and the functions $P_j(-\varphi)$ are defined by 5.5.3*. Here $\varphi(\underset{\sim}{x})$ stands for the normal density

$$\varphi(\underset{\sim}{x}) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}\underset{\sim}{x}' V^{-1} \underset{\sim}{x}\right),$$

---

* For the definition of $P_j(-\varphi)$, $P_n(-\Phi)$ etc., when the variance covariance matrix is not the identity matrix see the concluding remarks in section 5.5 of Chapter 5.

11 V being the variance-covariance matrix of $Z^{(1)}$.

The above theorem furnishes a uniform estimate of the difference $p_n(\underset{\sim}{z}) - q_n^{(s)}(\underset{\sim}{z})$. However, since our main interest lies in the estimation of the probabilities $P[Y^{(n)} \in A] = \sum_{\underset{\sim}{z} \in A} p_n(\underset{\sim}{z})$ it is necessary to study the behaviour of the quantity

$$\sum_{\underset{\sim}{z}} | p_n(\underset{\sim}{z}) - q_n^{(s)}(\underset{\sim}{z}) |$$

where $\underset{\sim}{z}$ runs over all the elements of the lattice (6.3.2). This is our object in the rest of the present section.

In what follows $\underset{\sim}{z}$ always denotes an element of (6.3.2) and $c_1, c_2, \ldots$ stand for constants independent of n but possibly depending on the distribution of $Z^{(1)}$.

Lemma 6.3.1. There exists a constant c_1 such that

$$(6.3.5) \qquad \sum_{|\underset{\sim}{z}|>\lambda} | q_n^{(s)}(\underset{\sim}{z}) | \leq c_1 \lambda^{(k-2)} \exp(-\lambda^2/8)$$

Proof. From the definition of the function $\tilde{P}_j(-\varphi)$ it is clear that it is of the form $g(\underset{\sim}{x})\,\varphi(\underset{\sim}{x})$ where $g(\underset{\sim}{x})$ is a multinomial in $x_1, \ldots, x_k$. Consequently it follows without difficulty that there exists a constant c_2 such that

$$(6.3.6) \qquad | \sum_{j=0}^{s-2} n^{-\frac{1}{2}j} \tilde{P}_j(-\varphi) | \leq c_2 \exp(-|\underset{\sim}{x}|^2/4)$$

If $\underset{\sim}{z} = (n\underset{\sim}{a} + \underset{\sim}{m})\, n^{-\frac{1}{2}}$ and $\underset{\sim}{z} \geq 0$, then it is easy to see that

$$n^{-\frac{1}{2}k}\, e^{-|\underset{\sim}{z}|^2/4} \leq \int_{\underset{\sim}{z} \leq \underset{\sim}{x} \leq \underset{\sim}{z}+n^{-\frac{1}{2}}} \exp\left(-\frac{1}{4}\,|\underset{\sim}{x}|^2\right) d\underset{\sim}{x}$$

with similar inequalities holding if $\underset{\sim}{z}$ lies in other regions such as $\underset{\sim}{z} \leq 0$, or $(z_1 < 0,\ z_j \geq 0 \text{ for } j \geq 2)$ etc. From this it follows at once that

$$(6.3.7) \qquad n^{-\frac{1}{2}k} \sum_{|\underset{\sim}{z}| > \lambda} \exp\left(-|\underset{\sim}{z}|^2/4\right) \leq c_3 \int_{|\underset{\sim}{x}| > \frac{1}{2}\lambda} \exp\left(-|\underset{\sim}{x}|^2/4\right) d\underset{\sim}{x}.$$

It can be shown by direct evaluation that

$$(6.3.8) \qquad \int_{|\underset{\sim}{x}| > \frac{1}{2}\lambda} \exp\left(-|\underset{\sim}{x}|^2/4\right) d\underset{\sim}{x} \leq c_4\, \lambda^{(k-2)} \exp\left(-\lambda^2/8\right).$$

Combining (6.3.6) - (6.3.8) we obtain the inequality (6.3.5). The lemma is thus proved.

Lemma 6.3.2. Let $\xi_1, \xi_2, \ldots,$ be a sequence of independent and identically distributed real valued random variables with mean values zero and $E\,|\xi_1|^s < \infty$ $(s \geq 3)$. Let $G_n(x)$ be the distribution function of $\eta_n = (\xi_1 + \ldots + \xi_n)/\sigma\sqrt{n}$. Then

$$|G_n(x) - \Phi(x)| \leq \frac{c}{1 + |x|^s}\, \frac{1}{n^{\frac{1}{2}(s-2)}}$$

for

$$|x| > ((1 + \delta)(s - 2) \log n)^{\frac{1}{2}}$$

where (1) $\Phi(x)$ is the normal distribution with mean zero, unit variance, (2) δ is any number such that $0 < \delta < 1$, and (3) c is a constant depending only on δ and the first s moments of ξ_1.

For the proof see Esseen [15], (Theorem 2, p. 73). As a consequence of this lemma we have

Lemma 6.3.3. Suppose that $\beta_s < \infty$. Then there exists a constant α depending on the moments of $\underset{\sim}{Z}^{(1)}$ such that for $\lambda > (\alpha \log n)^{\frac{1}{2}}$, we have

$$(6.3.9) \qquad \sum_{|\underset{\sim}{z}| > \lambda} p_n(\underset{\sim}{z}) \leq c_5 \frac{1}{(\log n)^s} \frac{1}{n^{\frac{1}{2}(s-2)}}$$

Proof. If $Y_j^{(n)}$ denotes the jth coordinate of $Y^{(n)}$ then

$$Y_j^{(n)} = (Z_j^{(1)} + \ldots + Z_j^{(n)}) / n^{\frac{1}{2}}$$

Let σ_i^2 denote the variance of $Z_j^{(1)}$. Then by Lemma 6.3.2 we have

$$(6.3.10) \qquad P[|Y_j^{(n)}| > \lambda \sigma_j] \leq \frac{c}{1 + |\lambda|^s} \frac{1}{n^{\frac{1}{2}(s-2)}}$$

$$+ \int_{|x| > \lambda} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2} x^2) dx$$

for $\lambda > [(1 + \delta)(s - 2) \log n]^{\frac{1}{2}}$. Since

(6.3.11) $$\int_{|x| > \lambda} \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right)dx \le \frac{1}{\sqrt{2\pi}} \frac{1}{\lambda} \exp\left(-\tfrac{1}{2}\lambda^2\right)$$

It follows from (6.3.10) and (6.3.11) that

(6.3.12) $$P[|Y_j^{(n)}| > \sigma_j \lambda] \le c_6 \frac{1}{(\log n)^s} \frac{1}{n^{\frac{1}{2}(s-2)}}$$

for all $\lambda > \left(\frac{3}{2}(s-2)\log n\right)^{\frac{1}{2}}$. Now

(6.3.13) $$\sum_{|\underset{\sim}{z}| > \lambda} p_n(\underset{\sim}{z}) = P\left[|Y^{(n)}| > \lambda\right]$$

$$\le \sum_{j=1}^{k} P\left[|Y_j^{(n)}| > \lambda k^{-\frac{1}{2}}\right].$$

Thus for $\lambda > \left(\frac{3}{2k}(s-2)\log n\right)^{\frac{1}{2}} \max_{1 \le j \le n} \sigma_j$, we have from (6.3.12) and (6.3.13) that

$$\sum_{|\underset{\sim}{z}| > \lambda} p_n(\underset{\sim}{z}) \le c_7 (\log n)^{-s} n^{-\frac{1}{2}(s-2)}$$

The lemma is thus proved.

<u>Theorem</u> 6.3.2. Suppose that $\beta_s < \infty$. Then

(6.3.14) $$\sum_{\underset{\sim}{z}} |p_n(\underset{\sim}{z}) - q_n^{(s)}(\underset{\sim}{z})| = o\left(\frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}}\right).$$

where $q_n^{(n)}(\underset{\sim}{x})$ is defined by (6.3.4).

15

Proof. From Theorem 6.3.1. it follows that

$$| p_n(\underset{\sim}{z}) - q_n^{(s)}(\underset{\sim}{z}) | \leq O\left(n^{-\frac{1}{2}(s+k-2)}\right)$$

uniformly in $\underset{\sim}{z}$. Consequently it follows that

$$(6.3.15) \quad \sum_{|\underset{\sim}{z}| < (\alpha \log n)^{\frac{1}{2}}} |p_n(z) - q_n^{(s)}(z)| = O\left\{(\log n)^{\frac{1}{2}k}\, n^{-\frac{1}{2}(s-2)}\right\}$$

since the number of lattice points inside the sphere $|z|$ $|\underset{\sim}{z}| < (\alpha \log n)^{\frac{1}{2}}$ is of the order $(n \log n)^{\frac{1}{2}k}$. Now we may suppose that the constant α defined in Lemma 6.3.3 is such that $\alpha^2 > 2(s-2)$. No loss of generality is involved in this assumption. Hence from Lemma 6.3.1 we have

$$(6.3.16) \quad \sum_{|\underset{\sim}{z}| > (\alpha \log n)^{\frac{1}{2}}} |q_n^{(s)}(\underset{\sim}{z})| = O\left(n^{-\frac{1}{2}(s-2)}\right).$$

Again by Lemma 6.3.3 it follows that

$$(6.3.17) \quad \sum_{|\underset{\sim}{z}| > (\alpha \log n)^{\frac{1}{2}}} p_n(\underset{\sim}{z}) \leq c_{10} (\log n)^{-s}\, n^{-\frac{1}{2}(s-2)}$$

Combining (6.3.15)-(6.3.17) we obtain the assertion of the theorem.

In the above theorem the error is shown to be of the order of $(\log n)^{\frac{1}{2}k}\, n^{-\frac{1}{2}(s-2)}$, where it is natural to expect that the logarithmic factor is superfluous. However it may be pointed

out that if we suppose that $\beta_{s+2} < \infty$ then the error in (6.3.14) is actually of the order $n^{-\frac{1}{2}(s-2)}$. This follows from the facts that

$$q_n^{s+1}(\underset{\sim}{x}) - q_n^{s}(\underset{\sim}{x}) = n^{-\frac{1}{2}(k+s-2)}\, \tilde{P}_{s-2}(-\varphi),$$

and

$$\sum_{\underset{\sim}{z}} |q_n^{s+1}(\underset{\sim}{z}) - q_n^{s}(\underset{\sim}{z})| = n^{-\frac{1}{2}(k+s-2)} \sum_{z} |\tilde{P}_{s-2}(-\varphi)|$$
$$= O(n^{-\frac{1}{2}(s-2)}).$$

Thus if $\beta_{s+2} < \infty$, then

$$(6.3.18) \qquad \sum_{\underset{\sim}{z}} |p_n(\underset{\sim}{z}) - q_n^{s}(\underset{\sim}{z})| = O(n^{-\frac{1}{2}(s-2)}).$$

In particular, if $\beta_5 < \infty$, we have

$$(6.3.19) \qquad \sum_{\underset{\sim}{z}} |p_n(\underset{\sim}{z}) - n^{-\frac{1}{2}k}\,[\,\varphi(\underset{\sim}{z}) + n^{-\frac{1}{2}}\,\tilde{P}_1(-\varphi)]| = O(n^{-1})$$

Another result in this direction is the following

Theorem 6.3.3. Suppose that $\beta_5 < \infty$.

Then

$$(6.3.20) \qquad \sum_{\underset{\sim}{z}} |q_n(\underset{\sim}{z}) - n^{-\frac{1}{2}k}\,\varphi(z)| = \frac{\lambda}{\sqrt{n}} + O\left(\frac{1}{n}\right)$$

where the summation extends over all elements of lattice 6.3.2, and the constant λ is given by

$$\lambda = \int | \tilde{P}_1 (-\varphi) | \, d\underset{\sim}{x} \tag{6.3.21}$$

Proof. From(6.3.19) we have that if I denotes the left hand side of the relation (6.3.20) then

$$I \leq n^{-\frac{1}{2}(k+1)} \sum_{\underset{\sim}{z}} | \tilde{P}_1 (-\varphi)| + 0 \left(\frac{1}{n}\right) \tag{6.3.22}$$

From the generalized Euler-Maclaurin sum formula (cf. Theorem B.3.2, of Appendix B) it is easily deduced that

$$n^{-\frac{1}{2}k} \sum_{\underset{\sim}{z}} |\tilde{P}_1 (-\varphi)| = \int |\tilde{P}_1(-\varphi)| \, d\underset{\sim}{x} + 0 \left(n^{-\frac{1}{2}}\right) \tag{6.3.23}$$

Also

$$n^{-\frac{1}{2}(k+1)} |\tilde{P}_1(-\varphi)| \leq | p_n(\underset{\sim}{z}) - n^{-\frac{1}{2}k}[\varphi + n^{-\frac{1}{2}}\tilde{P}_1(-\varphi)]| \tag{6.3.24}$$

$$+ |\bar{p}_n(\underset{\sim}{z}) - n^{-\frac{1}{2}k} \varphi (\underset{\sim}{z})|$$

From (6.3.19) and (6.3.24) it follows that

$$n^{-\frac{1}{2}(k+1)} \sum_{\underset{\sim}{z}} |\tilde{P}_1(-\varphi)| \leq I + 0 (n^{-1}) \tag{6.3.25}$$

Combining (6.3.22), (6.3.23) and (6.3.25) we obtain

$$I = n^{-\frac{1}{2}(k+1)} \sum_{\underset{\sim}{z}} |\tilde{P}_1(-\varphi)| + 0\ (n^{-1}) =$$

$$= \lambda\, n^{-\frac{1}{2}} + 0\ (n^{-1})$$

where λ is given by (6.3.21). The proof is thus complete.

In particular for real valued random variables we have the following

Theorem 6.3.4. Let $X_1, X_2, \ldots,$ be a sequence of independent real valued random variables with a common lattice distribution and such that $E\,X_i = 0$ and $E\,X_i^2 = \sigma^2$. Let $Y_n = (X_1 + \ldots + X_n)/\, n^{\frac{1}{2}}\sigma$. Then

$$(6.3.26) \qquad \sum_{z} |P[\,Y_n = z] - \frac{1}{\sqrt{2\pi n}\,\sigma} \exp\left(-\tfrac{1}{2} z^2\right)| = \frac{|\alpha_3|}{6\sigma^3} \cdot \frac{c}{\sqrt{2\pi n}} + 0\left(\frac{1}{n}\right)$$

where

$$c = \int_{-\infty}^{\infty} |3x - x^3| \exp\left(-\tfrac{1}{2} x^2\right) dx$$

$$= 2(1 + 4e^{-3/2}) < 0.251$$

where $\alpha_3 = E\,X_i^3$, and the summation (6.3.26) extends over all discontinuity points z of Y_n.

19 This follows from Theorem 6.3.3 and the fact that

$$P_1(-\varphi) = \frac{\alpha_3}{6\sigma^3} \frac{1}{\sqrt{2\pi}\,\sigma} \left(3\frac{x}{\sigma} - \frac{x^3}{\sigma^3}\right) \exp\left(-x^2/2\,\sigma^2\right).$$

Remarks. Theorems 6.3.2, 6.3.3 and 6.3.4 all seem to be new. A particular case of Theorem 6.3.4 has been obtained earlier by Prohorov (cf. [30]). It concerns the case when X_i is a random variable taking only the values 0 and 1 with probabilities q and p respectively. In this case, $\sigma^2 = pq$ and $\alpha_3 = E(X_i - p)^3 = pq\,(q - p)$. We thus have

$$\sum_r \left| \binom{n}{r} p^r q^{n-r} - \frac{1}{\sqrt{2\pi\, npq}} \exp\left(-\tfrac{1}{2} x_r^2\right) \right|$$

$$= \frac{|p - q|}{3\sqrt{2\pi}} \frac{(1 + 4e^{-3/2})}{(pqn)^{\frac{1}{2}}} + O\left(\frac{1}{n}\right)$$

where $x_r = (r - np)/(npq)^{\frac{1}{2}}$.

6.4. Asymptotic expansion of the probability $P_n(A)$ and estimation of the remainder term. An immediate consequence of Theorem 6.3.2 is that if

$$(6.4.1) \qquad Q_n^{(s)}(A) = \sum_{\underset{\sim}{z} \in A} q_n^{(s)}(\underset{\sim}{z})$$

where $\underset{\sim}{z}$ is an element of the lattice (6.3.2), then

$$(6.4.2) \qquad \sup_{A} |P_n(A) - Q_n^{(s)}(A)| = O\left(\frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}}\right)$$

where $P_n(A) = P[\, Y^{(n)} \varepsilon A]$. This follows from the fact that

$$|P_n(A) - Q_n^{(s)}(A)| \leq \sum_{\underset{\sim}{z}} |p_n(\underset{\sim}{z}) - q_n^{(s)}(\underset{\sim}{z})|$$

for all measurable sets A. The question now arises whether it would be possible to give a closed expression for $Q_n^{(s)}(A)$ at least for some important classes of sets such as the rectangles in R_k. This can indeed be achieved with the help of the generalized Euler-Maclaurin sum formula.

Let $\Phi(\underset{\sim}{x})$ denote the normal distribution in R_k with mean zero and variance-covariance matrix = V. i.e.

$$\Phi(\underset{\sim}{x}) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \int_{-\infty < \underset{\sim}{y} \leq \underset{\sim}{x}} \exp\left(-\tfrac{1}{2}\, \underset{\sim}{y}'\, V^{-1}\, \underset{\sim}{y}\right) d\underset{\sim}{y}$$

Let $\Psi_n^{(s)}$ be defined by

$$(6.4.3) \qquad \Psi_n^{(s)}(\underset{\sim}{x}) = \sum_{0}^{s-3} n^{-\frac{1}{2}s}\, \tilde{P}_j(-\Phi)$$

Let $S_1(t), S_2(t), \ldots, S_j(t) \ldots$ for real t be the sequence of periodic functions described in Appendix B. (cf. (B.1.1)). Let $T_j^{(s)}$ denote the operator defined as follows: For $s \geq 3$

$$(6.4.4) \qquad T_j^{(s)} = 1 - n^{-\frac{1}{2}} S_1[\, x_j\sqrt{n} - na_j\,]\ \frac{\partial}{\partial x_j} + \cdots$$

$$+ (-1)^{s-3}\, n^{-\frac{1}{2}(s-3)}\, S_{s-3}\,[x_j\sqrt{n} - na_j]\ \frac{\partial^{s-3}}{\partial x_j^{s-3}}\,.$$

Here the vector $\underset{\sim}{a} = (a_1, \ldots, a_k)$ is that determining the lattice (6.3.1) or (6.3.2). Let $\prod_n^{(s)}(x)$ be defined as follows:

$$(6.4.5) \qquad \prod_n^{(s)}(x) = \Big[\prod_{j=1}^{k} T_j^{(s)}\Big]\,\Psi_n^{(s)}(x)$$

where $\Psi_n^{(s)}$ is given by (6.4.3).

We then have

Theorem 6.4.1. Suppose $\beta_s < \infty$. Then

$$(6.4.6) \qquad P\,[\,Y^{(n)} \leq \underset{\sim}{x}\,] = \prod_n^{(s)}(\underset{\sim}{x}) + O\left\{\frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}}\right\}$$

In particular, if $\beta_{s+2} < \infty$, then

$$(6.4.7) \qquad P\,[\,Y^{(n)} \leq \underset{\sim}{x}\,] = \prod_n^{(s)}(\underset{\sim}{x}) + O\,(n^{-\frac{1}{2}(s-2)})$$

where in both (6.4.6) and (6.4.7) the estimations hold uniformly in $\underset{\sim}{x}$.

Proof. From Theorm 6.3.2 and (6.4.2) it follows that

22

$$(6.4.8)\quad P[\,Y^{(n)} \underset{\sim}{\leq} \underset{\sim}{x}] - \sum_{\underset{\sim}{z} \underset{\sim}{\leq} \underset{\sim}{x}} q_n^{(s)}(\underset{\sim}{x}) = o\left\{\frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}}\right\}$$

uniformly in $\underset{\sim}{x}$, where $q_n^{(s)}(\underset{\sim}{x})$ is defined by (6.3.4), i.e.

$$q_n^{(s)}(\underset{\sim}{x}) = \sum_{j=o}^{(s-3)} n^{-\frac{1}{2}j}\, \tilde{P}_j(-\varphi)$$

where

$$\varphi(\underset{\sim}{x}) = \frac{|V|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}} \exp(-\tfrac{1}{2}\, \underset{\sim}{x}\, V^{-1}\, \underset{\sim}{x}).$$

Consequently it follows from the definition of $\tilde{P}_j(-\varphi)$ that

$$(6.4.9)\quad \Psi_n^{(s)}(x) = \int_{-\infty < \underset{\sim}{y} \leq \underset{\sim}{x}} q_n^{(s)}(\underset{\sim}{y})\, d\underset{\sim}{y}$$

where $\Psi_n^{(s)}(x)$ is defined by (6.4.3). Thus it follows from Theorem (B.3.1) of Appendix B and (6.4.9) that

$$(6.4.10)\quad \sum_{\underset{\sim}{z} \underset{\sim}{\leq} \underset{\sim}{x}} q_n^{(s)}(\underset{\sim}{x}) = \prod_{j=1}^{k}\Big\{T_j^{(s)} + (-1)^{s-2} n^{-\frac{1}{2}(s-2)} \int_{-\infty}^{x_j} S_{s-3}[x_j/\sqrt{n} - na_j]\, \frac{\partial^{s-2}}{\partial x_j^{s-2}}\Big\} \Psi_n^{(s)}(x).$$

It is easily verified that the left hand side of (6.4.10) reduces to

$$(6.4.11) \qquad \prod_n^{(s)} (\underset{\sim}{x}) + 0\,(n^{-\frac{1}{2}(s-2)}).$$

By combining (6.4.8), (6.4.10) and (6.4.11) we obtain

$$P[\,Y^{(n)} \leq \underset{\sim}{x}] = \prod_n^{(s)} (\underset{\sim}{x}) + 0\left\{\frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}}\right\}.$$

uniformly in $\underset{\sim}{x}$. This proves (6.4.6). (6.4.7) is an immediate consequence of (6.4.6).

Theorem 6.4.2. Suppose that $\beta_5 < \infty$. Then

$$(6.4.12) \quad P[\,Y^{(n)} \leq \underset{\sim}{x}] = \Phi\,(\underset{\sim}{x}) + n^{-\frac{1}{2}}\,\tilde{P}_1(-\Phi) - n^{-\frac{1}{2}} \sum_{j=1}^{k} S_1[x_j\sqrt{n} - na_j]\,\frac{\partial \Phi}{\partial x_j} + 0\,(\frac{1}{n})$$

uniformly in $\underset{\sim}{x}$.

Proof. For $s = 5$ it is easily deduced from the definition 6.4.5. of $\prod_n^{(s)}$ that

$$\prod_n^{(s)} (\underset{\sim}{x}) = \Phi\,(\underset{\sim}{x}) + n^{-\frac{1}{2}}\,\tilde{P}_1\,(-\Phi\,) - n^{-\frac{1}{2}} \sum_{j=1}^{k} S_1[x_j\sqrt{n} - na_j]\,\frac{\partial \Phi}{\partial x_j} + 0\,(\frac{1}{n}).$$

Theorem 6.4.2 follows hence as an immediate consequence of Theorem 6.4.1.

Another easy consequence of Theorem 6.4.1. is the expansion in the case $k = 1$. In this case we have

Theorem 6.4.3. Let $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed random lattices variables with means zero and variances σ^2, and let the lattice be $[a + r: r = 0, \pm 1, \pm 2, \ldots]$. Let $F_n(x) = P[(X_1+\ldots+X_n) \leq x\sigma\sqrt{n}\,]$. If $E\,|X_1|^{s+2} < \infty$ then

$$F_n(x) = \Big\{1 - n^{-\frac{1}{2}} S_1[\,x\sigma\sqrt{n} - na]\ \frac{d}{dx} + \ldots + (-1)^{(s-3)} n^{-\frac{1}{2}(s-3)}\, S_{s-3}[2\sigma\sqrt{n} - na]\ \frac{d^{s-3}}{dx^{s-3}}\Big\}\Psi_n^{(s)}(x) + o\big(n^{-\frac{1}{2}(s-2)}\big).$$

uniformly in x, where $\Psi_n^{(s)}(x)$ is defined by (6.4.3).

Theorem (6.4.3) was obtained earlier by Esseen ([15], Theorem 4, p.61). His method for the one dimensional case is different from the present ones and is powerful enough to enable him to assume only the existence of $E|X_i|^s$. However, in the special case of $k=1$, it is possible to refine our method so that Esseen's theorem is obtained. This can be done as follows. First an improved version of Theorem 6.3.1 in the form

$$|z|^2\,|\,p_n(z) - q_n^{(s)}(z)\,| = o(n^{-\frac{1}{2}(s+k-2)})$$

where z ranges over the lattice, is obtained. This leads to the necessary improvements in all the succeeding theorems, and it becomes plain that the log n factor appearing in our preceding

estimation of the remainder is actually superfluous.

Finally we may remark that an analogue of Theorem 6.4.1. can be given for arbitrary measurable sets A. In fact we have

Theorem 6.4.4. Suppose $\beta_s < \infty$. Then

$$|P_n(A) - \int_A d\, T_n^{(s)}(\underset{\sim}{x})| = o\left\{\frac{(\log n)^{\frac{1}{2}k}}{n^{\frac{1}{2}(s-2)}}\right\}$$

uniformly for all Borel sets A. Thus in particular if $\beta_5 < \infty$

$$P_n(A) = \int_A dQ(\underset{\sim}{x}) + \frac{1}{\sqrt{n}} \int_A d\tilde{P}_1(-\underset{\sim}{D})\,\underset{\sim}{\Phi}(\underset{\sim}{x}) + o\left(\frac{1}{n}\right)$$

where

$$Q(\underset{\sim}{x}) = \prod_{j=1}^{k}\left\{1 - n^{-\frac{1}{2}} S_1[x_j/\sqrt{n} - na_j]\ \frac{\partial}{\partial x_j}\right\}\underset{\sim}{\Phi}(\underset{\sim}{x})$$

APPENDIX A

On Convex sets

1. Introduction. This appendix contains such background material and results concerning convex sets as are relevant to the thesis. Some of the results given here are well known and can be found in one of the following references. Eggleston, Convexity ([14]); Bonnessen and Fenchel, Theorie der Konvexen Körper [7]; and 'Seminar on convex sets', conducted at the Institute for Advanced Study, by P. T. Bateman and others ([3]), mimeographed).

2. Inradius and minimum width of a convex set. All sets considered here are subsets of the k-dimensional Euclidean space R_k. A subset $C \subset R_k$ is said to be convex provided that $x, y \in C \Rightarrow \alpha x + (1 - \alpha)y \in C$ for every $\alpha > 0$, $0 < \alpha < 1$. In a sense the structure of convex sets of R_k is completely described by the half-spaces that contain it. Here it may be recalled that a half-space is a set of the form

$$[x : \ell(x) < p]$$

where $\ell(x)$ is a linear functional and p a real number. Also a hyper-plane is a set of the form

$$[x : \ell(x) = p]$$

It is then obvious that if C is a convex set with empty interior then there exists a hyperplane H such that $C \subset H$.

We shall be mainly concerned only with convex bodies i.e. convex sets with non-empty interior. The following notation will be employed.

$$C^{\circ} = \text{interior of } C$$

$$\bar{C} = \text{closure of } C$$

It is then well known that C° and $\bar{C}$ are also convex if C is so.

By a direction vector $\underset{\sim}{\ell}$ we mean a vector $\underset{\sim}{\ell} = (\ell_1, .., \ell_k)$ such that $\sum \ell_i^2 = 1$. We denote the inner product of two vectors $\underset{\sim}{\ell}$ and $\underset{\sim}{x}$ in R_k by $\underset{\sim}{\ell}'.\underset{\sim}{x}$. Now suppose that C is a bounded convex set. Then perpendicular to each direction there exist two support hyperplanes H_1 and H_2 which bound C. i.e. if $H_1 = [\underset{\sim}{x}: \underset{\sim}{\ell}'.\underset{\sim}{x} = a]$ and $H_2 = [\underset{\sim}{x}: \underset{\sim}{\ell}'.\underset{\sim}{x} = b]$ then

$$C \subset [\underset{\sim}{x}: a \leq \underset{\sim}{\ell}'.\underset{\sim}{x} \leq b]$$

and further H_1 and H_2 both have non-empty intersection with the closure of C. Then the difference $b - a = |a - b|$ = is called <u>width</u> of the convex set C in the direction $\underset{\sim}{\ell}$, and is denoted by $w(\underset{\sim}{\ell}, C)$. One of the characteristics of a convex set, the <u>minimum width</u>, may then be defined by

$$w(C) = \inf_{\underset{\sim}{\ell}} w(\underset{\sim}{\ell}, C) \tag{A.2.1}$$

It is then clear that the minimum width is actually attained i.e. there exists an $\underset{\sim}{\ell}_o$ such that

$$w(C) = w(\underset{\sim}{\ell}_o, C).$$

In other words there exists numbers a and b ($a \leq b$) such that

$$C \subset [\underset{\sim}{x};\ a \leq \underset{\sim}{\ell}_0 \cdot \underset{\sim}{x} \leq b], \text{ and} \tag{A.2.2}$$

$$|b - a| \leq w(C).$$

Another characteristic of a convex set which plays an important role in our discussions is the in~radius. Let $S(\underset{\sim}{x}, \varrho)$ denote the sphere of radius ϱ with centre at $\underset{\sim}{x}$. Then the <u>in~radius</u> $r(C)$, may be defined as follows:

$$r(C) = \sup [\varrho:\ S(\underset{\sim}{x}, \varrho) \subset C,\ \text{for some } x]. \tag{A.2.3}$$

Obviously $r(C) = 0$ if and only if C has empty interior; $r(C) < \infty$ if C is bounded. Also it is obvious that $r(C) = r(\bar{C}) = r(C^0)$.

The main result of this section deals with certain inequalities between the in~radius and the minimal width of a convex set. For instance it is obvious that

$$w(C) \geq 2\, r(C)$$

A deeper result asserting an inequality in the reverse direction is given by the following well known theorem ([14], p 111).

<u>Theorem</u> A.1.1. For any bounded convex set C, the following always holds.

$$w(C) \leq 2\sqrt{k+2}\ \ r(C) \tag{A.2.4}$$

From the definition of the width $w(C)$ or the relation (A.2.2) it follows that theorem A.2.1. can be stated in the following form.

Theorem A.2.1'. Let C be a convex set such that $r(C) \leq \varepsilon$. Then there exists a direction vector $\underset{\sim}{l}$ such that

$$C \subset [x:\ a \leq \underset{\sim}{l}'.\ \underset{\sim}{x} \leq b\ ], \quad \text{and} \tag{A.2.5}$$

$$|b - a| \leq 2\sqrt{k+2}\ \varepsilon.$$

3. Gauge function of a convex set and its properties. In this section the properties of gauge functions associated with convex sets is developed.

Let a be an inner point of a convex set C. Then the gauge function* $F(x)$ of C with respect to a is defined as follows.

$$F(\underset{\sim}{x}) = \inf\ [\alpha;\ a + \alpha^{-1}(x - a)\ \varepsilon\ C] \tag{A.3.1}$$

Clearly $F(\underset{\sim}{x}) = 0$ if $\underset{\sim}{x} = \underset{\sim}{a}$ or if $\underset{\sim}{x} \neq \underset{\sim}{a}$ and there is no boundary point on the ray joining x and a in the direction $x - a$. If $x \neq a$ and there is a boundary point b on the ray joinint a and x then it is not difficult to verify that

$$F(\underset{\sim}{x}) = \frac{|x - a|}{|b - a|} \tag{A.3.2}$$

Thus $F(\underset{\sim}{x}) < 1$ if and only if $\underset{\sim}{x}\ \varepsilon\ C^{o}$ and $F(\underset{\sim}{x}) = 1$ if and only if $\underset{\sim}{x}$ ε Boundary of C and $F(\underset{\sim}{x}) > 1$ if only if $\underset{\sim}{x} \notin \bar{C}$.

---

* it is also sometimes known as the distance function, or the support function.

It may be remarked that the definition of a gauge function depends on two quantities, the inner point $\underset{\sim}{a}$ and the convex set C. When we want to draw attention to a specific inner point $\underset{\sim}{a}$ we will write the gauge function F as $F(\underset{\sim}{x}; \underset{\sim}{a}; C)$.

Theorem A.3.1. The gauge function $F(\underset{\sim}{x})$ of a convex set C with respect to the origin has the following properties.

(a) $F(\underset{\sim}{x}) \geq 0$ for each $\underset{\sim}{x}$

(b) $F(\lambda \underset{\sim}{x}) = \lambda F(\underset{\sim}{x})$ for $\lambda \geq 0$

(c) $F(\underset{\sim}{x} + \underset{\sim}{y}) \leq F(\underset{\sim}{x}) + F(\underset{\sim}{y})$

Proof. Properties (a) and (b) are obvious from the definition. If $F(\underset{\sim}{x}) = 0 = F(\underset{\sim}{y})$, then by (b), $2\alpha \underset{\sim}{x}$, $2\alpha \underset{\sim}{y} \in C$ for every $\alpha > 0$, and consequently $\alpha(\underset{\sim}{x} + \underset{\sim}{y}) = \frac{2\alpha x + 2\alpha y}{2} \in C$ for every α and therefore by definition $F(\underset{\sim}{x} + \underset{\sim}{y}) = 0$. If $F(\underset{\sim}{x}) \neq 0$ but $F(\underset{\sim}{y}) = 0$, $\underset{\sim}{x}/F(x)$ is a boundary point of C and $\underset{\sim}{y}/\alpha F(x)$ is an inner point of C for every positive α; thus

$$F\left\{ \frac{1}{1+\alpha} \quad \frac{x}{F(x)} + \frac{\alpha}{1+\alpha} \quad \frac{y}{\alpha F(x)} \right\} < 1.$$

$F(\underset{\sim}{x} + \underset{\sim}{y}) < (1 + \alpha) F(\underset{\sim}{x})$, and since α is arbitrary

$$F(\underset{\sim}{x} + \underset{\sim}{y}) \leq F(\underset{\sim}{x}) = F(\underset{\sim}{x}) + F(\underset{\sim}{y}).$$

If $F(\underset{\sim}{x}) = \alpha \neq 0$ and $F(\underset{\sim}{y}) = \beta \neq 0$, then $\underset{\sim}{x}/\alpha$ and $\underset{\sim}{y}/\beta$ are in $\bar{C}$, so that

$$F\left\{ \frac{\alpha}{\alpha+\beta} \quad \frac{\underset{\sim}{x}}{\alpha} + \frac{\beta}{\alpha+\beta} \quad \frac{\underset{\sim}{y}}{\beta} \right\} \leq 1$$

Theorem 2.2.1) the closure of the family $\mathcal{F}_m^{(0)}$ is compact in the u.c.c. topology. Actually the family $\mathcal{F}_m^{(0)}$ is closed in the u.c.c. topology as is shown below.

Suppose $F_1(x), \ldots F_n(x), \ldots$, is a sequence of functions belonging to $\mathcal{F}_m^{(0)}$ converging in u.c.c. to a function F. From the definition of

and

$$F(\underset{\sim}{x} + \underset{\sim}{y}) \leq \alpha + \beta = F(\underset{\sim}{x}) + F(\underset{\sim}{y}).$$

This completes the proof.

Theorem A.3.2. If $F(x)$ is any real valued function on R_k having properties (a), (b), and (c) of theorem A.3.1. then the set $\{ \underset{\sim}{x} ; F(\underset{\sim}{x}) < 1 \}$ is an open convex set including the origin and $F(x)$ is the gauge function of this convex set.

Proof. Since $F(\underset{\sim}{x})$ is obviously continuous it follows that $C = [\underset{\sim}{x}; F(\underset{\sim}{x}) < 1]$ is an open set. Again convexity trivially follows from properties (b) and (c) of Theorem A.3.1. Condition (b) actually implies that $F(\underset{\sim}{x})$ is the gauge function of the set C.

Theorem A.3.3. Let C be a convex set and let $\underset{\sim}{a} \in C^{\circ}$. Let $F(\underset{\sim}{x})$ be the gauge function of C with respect to $\underset{\sim}{a}$. Then (1) $S(\underset{\sim}{a}, \varrho) \subset C$ if and only if

$$F(\underset{\sim}{x}) \leq \frac{1}{\varrho} |\underset{\sim}{x} - \underset{\sim}{a}| \tag{A.3.3}$$

for each $\underset{\sim}{x}$. (2) $S(\underset{\sim}{a}, \varrho) \supset C$ if and only if

$$F(\underset{\sim}{x}) \geq \frac{1}{\varrho} |\underset{\sim}{x} - \underset{\sim}{a}| \tag{A.3.4}$$

for each $\underset{\sim}{x}$.

Proof. First suppose $S(\underset{\sim}{a}, \varrho) \subset C$. Let $\underset{\sim}{b}$ be the boundary point on the ray joining $\underset{\sim}{a}$ and $\underset{\sim}{x}$ in the direction $\underset{\sim}{x}-\underset{\sim}{a}$. Then it is clear that

$$|\underset{\sim}{b} - \underset{\sim}{a}| \geq \varrho.$$

But $F(\underset{\sim}{x}) = |\underset{\sim}{x} - \underset{\sim}{a}| \,/\, |\underset{\sim}{b} - \underset{\sim}{a}| \leq (1/\varrho)\, |\underset{\sim}{x} - \underset{\sim}{a}|$ which proves (A.3.3). Conversely if (A.3.3) is fulfilled it follows that

$$S(\underset{\sim}{a}, \varrho) = [\underset{\sim}{x};\ \frac{1}{\varrho}\, |\underset{\sim}{x} - \underset{\sim}{a}| < 1] \subset [\underset{\sim}{x}:\ F(\underset{\sim}{x}) < 1] = C^{\circ}$$

i.e. $S(\underset{\sim}{a}, \varrho) \subset C$.

For proof of the second part first suppose that $S(\underset{\sim}{a}, \varrho) \supset C$. Then as before, if $\underset{\sim}{x}$ is any point and $\underset{\sim}{b}$ the boundary point on the ray from $\underset{\sim}{a}$ to $\underset{\sim}{x}$ it clearly follows that

$$|\underset{\sim}{b} - \underset{\sim}{a}| \leq \varrho.$$

Consequently $F(\underset{\sim}{x}) \geq (1/\varrho)\, |\underset{\sim}{x} - \underset{\sim}{a}|$ which proves (A.3.4). The converse is proved as in the first part.

This completes the proof of the theorem.

Corollary to Theorem A.3.1. For any bounded convex set C there exists a point $\underset{\sim}{z} \in C^{\circ}$ such that if $F(x)$ is the gauge function of C with respect to $\underset{\sim}{z}$, then

$$F(\underset{\sim}{x}) \leq \frac{1}{r(C)}\, |\underset{\sim}{x} - \underset{\sim}{z}| \tag{A.3.5}$$

for all $\underset{\sim}{x}$.

Proof. From the definition of $r(C)$ there exists a sequence of points $x_n \in C$ and a sequence of positive numbers $r_n \uparrow r(C)$ such that

$$S(x_n, r_n) \subset C$$

Let $\underset{\sim}{z}$ be a limit point of x_n. Such a limit point obviously exists since C is bounded. Then it follows that $\underset{\sim}{z} \in C^{\circ}$ and that

$$S(z, r(C)) \subset C.$$

The result is then a corollary of Theorem A.3.3.

The main purpose of the above theorems here is to prove theorem A.3.4 below which is needed in chapter 3. We begin with a lemma.

Lemma A.3.1. Let C be any bounded convex set then there exists a point $\underset{\sim}{z}$ belonging to the interior of C such that if $F(x)$ is the gauge function of C with respect to $\underset{\sim}{z}$, then

$$|F(\underset{\sim}{x}) - F(\underset{\sim}{y})| \leq \frac{1}{r(C)} |\underset{\sim}{x} - \underset{\sim}{y}| \tag{A.3.6}$$

for all $\underset{\sim}{x}$ and $\underset{\sim}{y}$.

Proof. Let $\underset{\sim}{z} \in C^{o}$ be the point such that

$$S(\underset{\sim}{z}, r(C)) \subset C. \tag{A.3.7}$$

Then consider the function $G(\underset{\sim}{x}) = F(\underset{\sim}{x} + \underset{\sim}{z})$. It is easily verified that $G(\underset{\sim}{x})$ is the gauge function of the set $C - \underset{\sim}{z}$ with respect to the origin. Consequently by property (c) of Theorem A.3.1,

$$G(\underset{\sim}{x}) - G(\underset{\sim}{y}) \leq G(\underset{\sim}{x} - \underset{\sim}{y}),$$
$$G(\underset{\sim}{y}) - G(\underset{\sim}{x}) \leq G(\underset{\sim}{y} - \underset{\sim}{x}).$$

Thus

$$|G(\underset{\sim}{x}) - G(\underset{\sim}{y})| \leq \text{Max}\left\{G(\underset{\sim}{x} - \underset{\sim}{y}),\ G(\underset{\sim}{y} - \underset{\sim}{x})\right\}. \tag{A.3.8}$$

In terms of the function F, this means that

$$|F(\underset{\sim}{x}) - F(\underset{\sim}{y})| \leq \text{Max}\left\{F(\underset{\sim}{x} - \underset{\sim}{y} + \underset{\sim}{a}),\ F(\underset{\sim}{y} - \underset{\sim}{x} + \underset{\sim}{a})\right\}. \tag{A.3.9}$$

The last inequality (A.3.9) is deduced from (A.3.8) by replacing $\underset{\sim}{x}$, $\underset{\sim}{y}$ by $(\underset{\sim}{x} - \underset{\sim}{a})$, $(\underset{\sim}{y} - \underset{\sim}{a})$ respectively. If we now combine (A.3.9) and (A.3.5), it

yields the inequality (A.3.6).

The proof of the lemma is thus complete.

Theorem A.3.4. Let $\mathcal{F}_m^{(\delta)}$ be the class of all functions F which satisfy the following two conditions

i) F is a gauge function of a convex set contained in S(o,m)

ii) for each x and y

$$|F(x) - F(y)| \leq |x - y| / \delta$$

Then the class $\mathcal{F}_m^{(\delta)}$ is compact in the u.c.c. topology. Also if $\mathcal{C}_m^{(\delta)}$ is the class of all sets defined by

$$\mathcal{C}_m^{(\delta)} = [C;\ C \text{ - convex and } r(C) \geq \delta]$$

then each set $C \in \mathcal{C}_m^{(\delta)}$ has a gauge function $F \in \mathcal{F}_m^{(\delta)}$. Also if μ is a measure for which $\mathcal{C}_m^{(\delta)}$ is a continuity class then μF^{-1} is continuous (or nonatomic) for each $F \in \mathcal{F}_m^{(\delta)}$.

Proof. Let F be an arbitrary function $\in \mathcal{F}_m^{(\delta)}$. By (i) F is a gauge function of some convex set, say $C \subset S(o,m)$ with respect to some point $\underset{\sim}{z} \in C^o$. Then obviously $F(\underset{\sim}{z}) = 0$ and therefore by (ii),

$$F(\underset{\sim}{x}) \leq (|\underset{\sim}{x}| + |\underset{\sim}{z}|) / \delta \leq (|\underset{\sim}{x}| + m) / \delta. \tag{A.3.10}$$

Condition (ii) above together with (A.3.10) implies that (by Theorem 2.2.1) the closure of the family $\mathcal{F}_m^{(\delta)}$ is compact in the u.c.c. topology. Actually the family $\mathcal{F}_m^{(\delta)}$ is closed in the u.c.c. topology as is shown below.

Suppose $F_1(x), \ldots F_n(x), \ldots$, is a sequence of functions belonging to $\mathcal{F}_m^{(\delta)}$ converging in u.c.c. to a function F. From the definition of

Finally suppose μ is a measure for which each convex set is a continuity set. Let $F \in \mathcal{F}_m^{(\delta)}$. Since $F(x)$ is continuous the set

$$[\underset{\sim}{x} : F(\underset{\sim}{x}) = a]$$

is the boundary of the convex set

$$[\underset{\sim}{x} : F(\underset{\sim}{x}) \leq a]$$

and therefore it follows that μF^{-1} is nonatomic.

This completes the proof of the theorem.

4. Blaschke's theorem. This is one of the important classical theorems on convex sets. For the sake of completeness we present here a simple proof of the above theorem. (See [2], p 41 ; [14], p 64).

Let A and B be any two closed sets. Then the distance d(A, B) is defined as follows.

$$(A.4.1) \qquad d(A, B) = \inf [\varrho;\ S(A, \varrho) \supset B,\ \ S(B, \varrho) \supset A]$$

where $S(A, \varrho)$ $(S(B, \varrho))$ is the union of all spheres of radius ϱ with centres in A(B). It is easy to verify that d(A,B) is actually a distance function i.e. it has the following properties

(i) $d(A, B) \geq 0$, $= 0$ iff $A = B$

(ii) $d(A, B) = d(B, A)$

(iii) $d(A, B) \leq d(A, C) + d(B, C)$

Theorem A.4.1 (Blaschke) Let K be any cube. Then the class of all closed convex sets contained in K together with the distance d(A,B) is a compact metric space.

the family $\mathcal{F}_m^{(\delta)}$ we may suppose that F_n is the gauge function of a convex set $C_n \subset S(o, m)$ with respect to a point $\underset{\sim}{z}_n \in C_n^o$. Since $\underset{\sim}{z}_n$ is a bounded sequence we may suppose without loss of generality that

$$\underset{\sim}{z}_n \to \underset{\sim}{z} \in S(o, m).$$

Now consider the sequence $G_n(\underset{\sim}{x})$ of functions defined by

$$G_n(\underset{\sim}{x}) = F_n(\underset{\sim}{x} + \underset{\sim}{z}_n), \qquad n = 1, 2, \ldots .$$

Then clearly

$$G_n(\underset{\sim}{x}) \to G(\underset{\sim}{x}) = F(\underset{\sim}{x} + \underset{\sim}{z})$$

for each $\underset{\sim}{x}$. Since each $G_n(\underset{\sim}{x})$ satisfies conditions (a), (b) and (c) of Theorem A.3.1, the same is true of the function $G(\underset{\sim}{x})$. By Theorem A.3.2 $G(\underset{\sim}{x})$ and hence $F(\underset{\sim}{x})$ are both gauge functions. Let $C = [\underset{\sim}{x}: F(x) \leq 1]$ then $F(\underset{\sim}{x})$ is the gauge function of C with respect to $\underset{\sim}{z}$. It only remains to show that $C \subset S(o, m)$. Obviously

$$C \subset \text{limsup } C_n$$

and since each $C_n \subset S(o, m)$ it follows that $C \subset S(o, m)$. Thus the family $\mathcal{F}_m^{(\delta)}$ is compact in the u.c.c. topology.

Now suppose $C \in \mathcal{C}_m^{(\delta)}$. Then by lemma A.3.1. there exists a gauge function of C satisfying the condition

$$|F(\underset{\sim}{x}) - F(\underset{\sim}{y})| < |\underset{\sim}{x} - \underset{\sim}{y}| / r(C)$$
$$\leq |\underset{\sim}{x} - \underset{\sim}{y}| / \delta$$

since $r(C) \geq \delta$. Therefore $F \in \mathcal{F}_m^{(\delta)}$.

Proof. The proof will be carried out in three steps.

(a) If C_n is a sequence of closed convex sets and C is a closed set such that $d(C_n, C) \to 0$ then C is also convex. For suppose $\underset{\sim}{x}, \underset{\sim}{y} \in C$. Then if $\varepsilon > 0$ is arbitrary, it follows that there exists an n_o such that

$$C_{n_o} \subset S(C, \varepsilon); \quad C \subset S(C_{n_o}, \varepsilon) \tag{A.4.2}$$

Therefore $\underset{\sim}{x}, \underset{\sim}{y} \in S(C_{n_o}, \varepsilon)$ and since this set is convex, $\alpha \underset{\sim}{x} + (1 - \alpha)\underset{\sim}{y} \in S(C_{n_o}, \varepsilon)$. Also from A.4.2

$$S(C_{n_o}, \varepsilon) \subset S(C, 2\varepsilon),$$

so that $\alpha \underset{\sim}{x} + (1 - \alpha)\underset{\sim}{y} \in S(C, 2\varepsilon)$ for every $\varepsilon > 0$. Since C is closed this implies that $\alpha \underset{\sim}{x} + (1 - \alpha)\underset{\sim}{y} \in C$.

(b) If C_n is a sequence of closed and bounded sets such that $C_n \supset C_{n+1}$ for every n, then

$$d(C_n, C) \to 0$$

where $C = \bigcap_1^\infty C_n$. For suppose the contrary. Then since $d(C_n, C)$ is decreasing there would exist an $\varepsilon > 0$ such that

$$d(C_n, C) > \varepsilon \quad \text{for all } n.$$

Let $D_n = C_n - C_n \cap S(C, \varepsilon)$. Then D_n is closed and $D_n \supset D_{n+1}$ and D_n is non empty for each n. Hence

$$D = \bigcap_1^\infty D_n$$

is non empty and obviously $D \subset C$. Again from the definition of D_n it follows that $D_n \cap S(C, \varepsilon) = \emptyset$ for each n, i.e. $D \cap S(C, \varepsilon) = \emptyset$ which contradicts the fact that D is a non empty subset of C. This proves the statement (b).

The final step consists in the proof of the theorem itself. We may suppose without loss of generality that K is the unit cube. Let $\{A\}$ be an infinite collection of closed sets contained in K. By ith subdivision split cube K into 2^{ik} closed little cubes of side length 2^{-i}. Consider the set of all aggregates of these little cubes. The number of possible aggregates is finite $2^{2^{ik}}$. Now associate with each A of our infinite system $\{A\}$ the aggregate of all little cubes which have points in common with A. For each fixed i, at least one such aggregate occurs infinitely often.

Let $\{A_{1\nu}\}$ be an infinite sequence of our system such that each one of the sets $A_{1\nu}$ is associated with same aggregate C_1 in the first subdivision. Then let $\{A_{2\nu}\}$ be a subsequence of $\{A_{1\nu}\}$ each element of which is associated the same aggregate C_2 in the second subdivision and so on. Then obviously $A_{j\nu} \subset C_j$ and

$$d(A_{j\nu}, C_j) \leq \frac{\sqrt{k}}{2^j}.$$

Consequently for the infinite sequence $\{A_{jj}\}$ we have

$$d(A_{jj}, C_j) \leq \frac{k}{2^j} \to 0 \qquad (A.4.3)$$

as $j \to \infty$. Now C_j is a decreasing sequence of closed sets and if $C = \bigcap_1^\infty C_j$ it follows from (b) that

(A.4.4) $$\lim_{j \to \infty} d(C_j, C) = 0 .$$

Combining (A.4.3) and (A.4.4) one obtains that

$$\lim_{j \to \infty} d(A_{jj}, C) = 0 .$$

Finally if each one of sets $A \in \{A\}$ is convex it follows from (a) that C is also a convex set.

This completes the proof.

5. Convex sets and the Normal distribution in R_k. In this section we obtain estimates for the value of the Normal measure for certain convex sets.

Let Φ denote the Normal distribution in R_k with mean zero and variance covariance matrix equal to the identity matrix. i.e.

$$\Phi(A) = \int_A d\Phi = \int_A \left(\frac{1}{\sqrt{2\pi}}\right)^k \exp\left(-\tfrac{1}{2} \sum_{i=1}^{k} x_i^2\right) d\underset{\sim}{x}$$

Lemma A.5.1. Let C be any convex set for which $r(C) \leq \varepsilon$. Then

(A.5.1) $$\Phi(C) \leq \varepsilon \sqrt{\frac{2k+4}{\pi}}$$

Proof. By theorem A.2.1. (see also A.2.1') it follows that there exists a direction vector $\underset{\sim}{\ell}$ (i.e $\sum \ell_i^2 = 1$) such that

$$C \subset [\underset{\sim}{x}: \quad a \leq \underset{\sim}{\ell}' \underset{\sim}{x} \leq b]$$

where

$$|b - a| \leq 2(\sqrt{k+2})\,\varepsilon.$$

Now since the distribution of the function $\underset{\sim}{l}'\underset{\sim}{x}$ under the Normal distribution $\underset{\sim}{\Phi}$ is the Normal distribution with the density (on the real line)

$$\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad (-\infty < x < +\infty)$$

it follows that

$$\underset{\sim}{\Phi}(A) \leq \underset{\sim}{\Phi}[\underset{\sim}{x} : a \leq \underset{\sim}{l}'\underset{\sim}{x} \leq b]$$

$$\leq \frac{|b-a|}{\sqrt{2\pi}} \leq \varepsilon\sqrt{\frac{2(k+2)}{\pi}}$$

This proves the lemma.

Let us now introduce the following notation. For any $\alpha > 0$, let A^{α} denote the convex set $S(A, \alpha) = [\,\cup\, S(x, \alpha),\ x \in A]$. Also let $A^{-\alpha}$ denote the set

$$A^{-\alpha} = [\,\underset{\sim}{x}:\quad S(x, \alpha) \subset A]$$

It is clear that $A^{-\alpha} \neq \emptyset$ if $r(A) > \alpha$. It is also not difficult to verify that in that case $A^{-\alpha}$ is convex and has non-empty interior.

Lemma A.5.2. For any convex set C

$$\underset{\sim}{\Phi}(C^{\varepsilon} - C) \leq M\,\varepsilon \tag{A.5.2}$$

where M is a universal constant not depending upon C. If in addition $r(C) > \varepsilon$, so that $C^{-\varepsilon}$ is non-empty, then,

17

$$\text{(A.5.3)} \qquad \oint (C - C^{-\varepsilon}) \leq M\,\varepsilon.$$

Proof. It is not difficult to see that

$$\oint (C^{\varepsilon} - C) = \frac{1}{(2\pi)^{k/2}} \int_{C^{\varepsilon}-C} \exp\left(-\tfrac{1}{2}\textstyle\sum x_i^2\right) dx_1 \ldots dx_k$$

$$\text{(A.5.4)} \qquad = \frac{1}{(2\pi)^{k/2}} \int_0^{\varepsilon} d\alpha \int_{A^{\alpha}} \exp\left(-\tfrac{1}{2}\textstyle\sum x_i^2\right) dS$$

where the inner integral in the second equation is the surface integral over the surface of the convex set A^{α}. Similarly

$$\text{(A.5.5)} \qquad \oint (C - C^{-\varepsilon}) = \frac{1}{(2\pi)^{k/2}} \int_0^{\varepsilon} d\alpha \int_{A^{-\alpha}} \exp\left(-\tfrac{1}{2}\textstyle\sum x_i^2\right) dS$$

the inner integral representing as before the surface integral over the surface of the convex set $A^{-\alpha}$. From (A.5.4) and (A.5.5) it follows that for any convex set C

$$\text{(A.5.6)} \qquad \oint (C^{\varepsilon} - C) \leq \varepsilon \left(\frac{1}{2\pi}\right)^{k/2} \sup_S \int_S \exp\left(-\tfrac{1}{2}\textstyle\sum x_i^2\right) dS$$

$$\oint (C - C^{-\varepsilon}) \leq \varepsilon \left(\frac{1}{2\pi}\right)^{k/2} \sup_S \int_S \exp\left(-\tfrac{1}{2}\textstyle\sum x_i^2\right) dS$$

where the supremum in (A.5.6) is taken over all convex surfaces. Our next task is to show that this supremum is a finite constant.

First of all suppose that the origin is in the interior of the surface S. Let U denote the surface of the unit sphere (i.e. a sphere with centre at the origin and radius unity). For any point $\underset{\sim}{u}$ on the

surface U, let $\varrho(\underset{\sim}{u})$ denote the length of the segment from the origin to the point of the surface which lies on the ray joining the origin and the point $\underset{\sim}{u}$. Then it is elementary to show that

$$\int_S \exp\left(-\tfrac{1}{2}\sum x_i^2\right) dS = \int_U \exp\left(-\tfrac{1}{2}\varrho^2(\underset{\sim}{u})\right).\ \varrho(\underset{\sim}{u})du$$

where both integrals are the surface integrals respectively over the surfaces S and U. Since $|x|e^{-\frac{1}{2}x^2} \leq \sqrt{e}$ for all real x it follows that

$$\int_S \exp\left(-\tfrac{1}{2}\sum x_i^2\right) dS \leq \sqrt{e}\ L(U) = M$$

where $L(U)$ denotes the surface area of the unit sphere U.

Now suppose that the origin is not in the interior of the surface S. Consider the cone with vertex at the origin and containing surface S. It is clear that the cone divides the surface S into two portions S_1, S_2 say. The above argument can then be applied to each one of the integrals over S_1 and S_2 separately. Thus one obtains that in every case the inequality

$$\int_S \exp\left(-\tfrac{1}{2}\sum x_i^2\right) dS \leq M$$

holds, where M is a finite constant. By continuity of the integral as a function of the surface S it follows that the same result is valid even if the origin is on the surface S.

The proof is thus complete.

6. Convex sets and Fourier transforms. Let C be any bounded convex set with non-empty interior. Then its Fourier transform $q(\underset{\sim}{t}, C)$ is defined by

$$q(\underset{\sim}{t}, C) = \int_C \exp\,(i\, \underset{\sim}{t}'.\underset{\sim}{x})\, d\underset{\sim}{x} .$$

It follows from the multidimensional analogue of the Riemann Lebesgue lemma that

$$\text{(A.6.1)} \qquad \lim_{|\underset{\sim}{t}| \to \infty} q(\underset{\sim}{t}, C) = 0$$

where $|\underset{\sim}{t}| = (\sum_{i=1}^{k} t_i^2)^{\frac{1}{2}}$. The problem considered in this section is the rapidity of convergence in A.6.1.

Theorem A.6.1. Let C be any convex set $\subset S(0, m)$. Then

$$\text{(A.6.1)} \qquad |q(\underset{\sim}{t}, C)| \leq \frac{4}{|\underset{\sim}{t}|} \cdot \frac{\pi^{(k-1)/2}\, m^{k-1}}{\Gamma((k+1)/2)} .$$

Proof. For any vector $\underset{\sim}{t}$ and real number τ, let $H(\underset{\sim}{t}, \tau)$ denote the hyperplane

$$H(\underset{\sim}{t}, \tau) = [\, \underset{\sim}{x};\quad \underset{\sim}{t}'.\underset{\sim}{x} = \tau\, |\underset{\sim}{t}| \,] .$$

For fixed $\underset{\sim}{t}$, $H(\underset{\sim}{t}, \tau)$ are the hyperplanes which are perpendicular to the direction $\underset{\sim}{t}$. Let $H(\underset{\sim}{t}, \tau_1)$ and $H(\underset{\sim}{t}, \tau_2)$ with $\tau_1 \leq \tau_2$ be the two supporting hyperplanes of the convex set, perpendicular to the direction $\underset{\sim}{t}$. Since the convex set is bounded such finite τ_1, τ_2 exist.

It is then clear that

$$H(\underset{\sim}{t}, \tau) \cap C = \emptyset \quad \text{unless} \quad \tau_1 \leq \tau \leq \tau_2$$

Now let

$$p(\underset{\sim}{t}, \tau) = L_{k-1} \left[H(\underset{\sim}{t}, \tau) \cap C \right]$$

where L_{k-1} (A) denotes the (k-1) dimensional Lebesgue measure on the hyperplane $H(\underset{\sim}{t}, \tau)$. Then

$$(A.6.2) \qquad q(\underset{\sim}{t}, C) = \int_C \exp (i \, \underset{\sim}{t}' \cdot \underset{\sim}{x}) \, d\underset{\sim}{x}$$

$$= \int_{\tau_1}^{\tau_2} \exp (i \, |\underset{\sim}{t}| \tau) d\tau \cdot \int_{H(\underset{\sim}{t},\tau) \cap C} d L_{k-1}$$

$$= \int_{\tau_1}^{\tau_2} \exp (i \, \tau |\underset{\sim}{t}|) \, p \, (\underset{\sim}{t}, \tau) d\tau \, .$$

Now consider the function $p(\underset{\sim}{t}, \tau)$ in the interval $\tau_1 \leq \tau \leq \tau_2$. Obviously $p(\underset{\sim}{t}, \tau)$ is continuous and let τ_o ($\tau_1 \leq \tau_o \leq \tau_2$) be the point at which it attains its maximum. Then it is easily shown that $p(\underset{\sim}{t}, \tau)$ as a function of τ is nondecreasing from τ_1 to τ_o and nonincreasing from τ_o to τ_2 . Thus $p(\underset{\sim}{t}, \tau)$ is a function of bounded variation and

$$(A.6.3) \qquad \text{Variation of } p(\underset{\sim}{t}, \tau) \leq 2 \underset{\tau_1 \leq \tau \leq \tau_2}{\text{Max}} \; p(\underset{\sim}{t}, \tau) .$$

Since $p(\underset{\sim}{t}, \tau)$ is of bounded variation it follows from (A.6.2) by integration by parts, that

$$q(\underset{\sim}{t}, C) = \frac{e^{i\tau_2|\underset{\sim}{t}|}\, p(\underset{\sim}{t}, \tau_2) - e^{i\tau_1|\underset{\sim}{t}|}\, p(\underset{\sim}{t}, \tau_1)}{i\,|\underset{\sim}{t}|}$$

$$+ \frac{1}{i|\underset{\sim}{t}|} \int e^{i\tau|\underset{\sim}{t}|}\, dp(\underset{\sim}{t}, \tau)$$

i.e

$$|q(\underset{\sim}{t}, C)| \leq \frac{p(\underset{\sim}{t}, \tau_1) + p(\underset{\sim}{t}, \tau_2)}{|\underset{\sim}{t}|} + \frac{1}{|\underset{\sim}{t}|} \operatorname{Var}[p(\underset{\sim}{t}, \tau)].$$

Or from (A.6.3),

$$(A.6.4) \qquad |q(\underset{\sim}{t}, C)| \leq \frac{4}{|\underset{\sim}{t}|} \operatorname*{Max}_{\tau_1 \leq \tau \leq \tau_2} p(\underset{\sim}{t}, \tau).$$

Since $C \subset S(0, m)$ it is easy to see that

$$(A.6.5)' \qquad p(\underset{\sim}{t}, \tau) \leq \operatorname*{Max}_{\tau} L_{k-1}(H(\underset{\sim}{t}, \tau) \cap S(o, m))$$

$$\leq \frac{\pi^{(k-1)/2}\, m^{k-1}}{\Gamma((1+k)/2)}$$

The inequality (A.6.1) is then an immediate consequence of (A.6.4) and (A.6.5). This completes the proof.

Remarks. A more precise estimate of the rapidity of decrease of $q(\underset{\sim}{t}, C)$ could be obtained for certain special kinds of convex sets. We give two such examples.

First of all suppose C is a sphere with centre at the origin and radius r. i.e. $C = S(o, r)$. Then it is known (cf. Bochner [6], §43) that

$$q(\underset{\sim}{t}, C) = \left(\frac{2\pi r}{|\underset{\sim}{t}|}\right)^{k/2} J_{k/2}(r\,|\underset{\sim}{t}|),$$

$J_{k/2}(z)$ denoting the Bessel function of order $k/2$. Since the order of decrease of Bessel function is known i.e.

$$|J_{k/2}(z)| \leq \frac{M}{\sqrt{z}}$$

where M is a constant (see Watson [39], p 199) it follows that for $C = S(o, r)$

$$|q(\underset{\sim}{t}, C)| \leq M \frac{r^{(k-1)/2}}{|\underset{\sim}{t}|^{(k+1)/2}} \tag{A.6.6}$$

M being a constant depending only on k. The special nature of the convex set C is the cause of the much greater rapidity of decrease in this case. An estimate similar to (A.6.6) could also be obtained in the case of a convex sets C generated by a quadratic form. Let T be a positive definite matrix and let

$$Q(\underset{\sim}{x}) = \underset{\sim}{x}' T \underset{\sim}{x}.$$

Then consider convex sets C of the form

$$C = [\underset{\sim}{x}: \quad Q(\underset{\sim}{x}) \leq r\,].$$

By using a suitable linear transformation it could be deduced from (A.6.6) that in this case also

$$|q(\underset{\sim}{t}, C)| \leq M \frac{r^{(k-1)/2}}{|\underset{\sim}{t}|^{(k+1)/2}}$$

where the constant M depends only on the norm of T and T^{-1}.

22

As a second example consider the case when the convex set is a rectangle i.e.

$$C = [\underset{\sim}{x} : a_i \leq x_i \leq b_i; \quad i = 1, 2, \ldots, k]$$

In this case

$$q(\underset{\sim}{t}, C) = \prod_{j=1}^{k} \frac{e^{ib_j t_j} - e^{ia_j t_j}}{i t_j}$$

Thus if $t_{j_1}, t_{j_2}, \ldots, t_{j_r} \to \infty$ and the rest remain bounded then the order is

$$|q(\underset{\sim}{t}, C)| \leq \frac{M}{|t_{j_1} \ldots t_{j_r}|}$$

Incidentally this example is an extreme case in the class of all convex sets and actually shows that the general estimate of Theorem A.6.1 cannot be improved to $|\underset{\sim}{t}|^{-(1+\delta)}$ for any positive δ.

APPENDIX B

Generalization of the Euler-Maclaurin Sum Formula to Functions of Several Variables.

Let $f(x)$ be a function of the real variable x on the interval $0 \leq x \leq n$ with continuous differential coefficient $f'(x)$ Then a closed expression for the sum $\sum_{j=0}^{n} f(j)$ may be given as follows:

$$\sum_{j=0}^{n} f(j) = \int_0^n f(x)\,dx + \tfrac{1}{2}\,(f(0) + f(n)) + \int_0^n (x - [\,x\,] - \tfrac{1}{2})\, f'(x)\,dx$$

where $[\,x\,]$ denotes the least integer $\leq x$. This is the Euler-Maclaurin sum formula in its simplest form (cf., i.e., Knopp [23], Chapter 14). In this appendix our object is develop similar summation proceedures for functions on R_k, $k \geq 1$.

B.1. The functions $S_j(t)$. We now introduce a sequence of functions $S_j(t)$ $(j = 0, 1, 2, \ldots)$ which play an important role in what follows.

Let

$$S_0(t) = 1$$

$$(B.1.1.) \qquad S_j(t) = \begin{cases} (-1)^{\frac{1}{2}j-1} \sum_{n=1}^{\infty} \frac{2\cos(2n\pi t)}{(2n\pi)^j}; & j \text{ even} \\ (-1)^{\frac{1}{2}(j-1)} \sum_{n=1}^{\infty} \frac{2\sin(2n\pi t)}{(2n\pi)^j}; & j \text{ odd} \end{cases}$$

Then it is evident that $S_j(t)$ is a periodic function of t with period unity. Also

$$S_1(t) = t - [\,t\,] - \frac{1}{2}$$

is a right continuous function whose discontinuities are only at the points $0, \pm 1, \pm 2, \ldots$; at these points it has a jump of magnitude -1. For $j \geq 2$, $S_j(t)$ is absolutely continuous, and for all t

$$(B.1.2) \qquad S_j'(t) = \frac{d}{dt} S_j(t) = S_{j-1}(t).$$

It should be noted that (B.1.2) is valid for $j = 1$ as well, if we restrict t to intervals, $n < t < n+1$, where n is an integer, positive, negative or zero. Actually, for $0 < t < 1$, $S_j(t)$ is a polynomial in t. In fact for $0 < t < 1$

$$S_j(t) = \frac{1}{j!} \sum_{r=0}^{j} \binom{j}{r} B_r t^{n-r}$$

where $B_1, B_2, \ldots$ are the Bernoulli numbers. Further

$$S_{2j+1}(0) = S_{2j+1}(n) = 0$$

(B.1.3)

$$S_{2j}(n) = S_{2j}(0) = B_{2j}/(2j)!$$

These properties of the functions S_j are well known and may be found in the treatise of Knopp referred to.

B.2. The Euler-Maclaurin sum formula for functions on R_1. In this section we present the classical result for functions on R_1 in a form which is capable of generalisation to functions on R_k. Let $f(x)$ be a function of the real variable x possessing derivatives of all orders. We suppose that

$$1) \quad \lim_{|x| \to \infty} |f^{(r)}(x)| = 0$$

(B.2.1)

$$2) \quad \int |f^{(r)}(x)| \, dx < \infty$$

for $r = 0, 1, 2, \ldots$ where

$$f^{(r)}(x) = \frac{d^r f}{dx^r}$$

Then we have the following theorem.

Theorem B.2.1. For any measurable set A and any $r = 1, 2, \ldots$

$$\text{(B.2.2)} \qquad \sum_{n \in A} f(n) = \int_A d\,F_r(x)$$

where

$$\text{(B.2.3)} \qquad F_r(x) = \Big[\, 1 - S_1(x)\frac{d}{dx} + \ldots + (-1)^r S_r(x)\,\frac{d^r}{dx^r} + (-1)^{r+1} \int_{-\infty}^{x} S_r(x)\,\frac{d^{r+1}}{dx^{r+1}}\,\Big]\, F(x)$$

and

$$\text{(B.2.4)} \qquad F(x) = \int_{-\infty}^{x} f(x)\,dx$$

Proof. It is clear from the definition of $F_r(x)$ that it is a function of bounded variation. Also since each one of functions $S_j(x)$ is absolutely continuous with a continuous differential coefficient (cf. B.1.2) in the interval $n < x < n+1$, n being an integer, it follows that $F_r(x)$ is also absolutely continuous for $n < x < n+1$. By using the property (B.1.2) it is easily verified that

$$\frac{d}{dx}\,F_r(x) = 0$$

for $n < x < n+1$, i.e. $F_r(x)$ remains constant in these intervals.

Now, $S_1(x)$ is discontinuous and has a jump of magnitude -1 at the points $x = n$, and hence $F_r(x)$ also has jumps at the points $x = n$, the value of the jump being $= f(n)$. Thus $F_r(x)$ is a step function with jumps of magnitude $f(n)$ at the points $x = n$. This proves the theorem.

Corollary 1. For every real x

$$(B.2.5) \qquad \sum_{-\infty < n \leq x} f(n) = F_r(x)$$

where $F_r(x)$ is defined by (B.2.3).

By a linear change of variable in (B.2.5) the following result is easily deduced. Let a and h be arbitrary real numbers. Then

$$(B.2.6) \qquad h \sum_{-\infty < a + hn \leq x} f(a + hn) = G_r(x)$$

where

$$(B.2.7) \qquad G_r(x) = F(x) - hS_1[(x - a)/h]\, F^{(1)}(x) + \ldots$$
$$+ (-1)^r h^r S_r[(x - a)/h]\, F^{(r)}(x) +$$
$$(-1)^{r+1} h^{r+1} \int_{-\infty}^{x} S_r[(x - a)/h]\, F^{(r+1)}(x)dx.$$

Here $F^{(j)}(x)$ denotes the j-th differential coefficient of $F(x)$.

B.3. Generalization to functions of several variables. Let R_k denote the k-dimensional Euclidean spaces. Vectors belonging to R_k will be denoted by $\underset{\sim}{x} = (x_1, \ldots, x_k)$. If $\underset{\sim}{x} = (x_1, \ldots, x_k)$ then $|\underset{\sim}{x}|^2 = x_1^2 + \ldots + x_k^2$. By an integer vector we mean a vector $\underset{\sim}{n} = (n_1, \ldots, n_k)$ with integral coordinates n_j; positive, negative or zero. For two vectors $\underset{\sim}{x} = (x_1, \ldots, x_k)$ and $\underset{\sim}{y} = (y_1, \ldots, y_k)$ we write $\underset{\sim}{x} \leq \underset{\sim}{y}$ if $x_j \leq y_j$ for $j = 1, 2, \ldots, k$.

Let $f(\underset{\sim}{x}) = f(x_1, \ldots, x_k)$ be a function on R_k possessing partial differential coefficients of all orders. We suppose in addition that

(B.3.1)

$$1) \quad \lim_{|\underset{\sim}{x}| \to \infty} \left| \frac{\delta^{s_1+\ldots+s_k}}{\delta x_1^{s_1} \ldots \delta x_k^{s_k}} f \right| = 0 < \infty$$

$$2) \quad \int_{R_k} \left| \frac{\delta^{s_1+\ldots+s_k}}{\delta x_1^{s_1} \ldots \delta x_k^{s_k}} f \right| d\underset{\sim}{x} < \infty$$

for all positive integers $s_1, \ldots, s_k$.

Let

(B.3.2) $$F(\underset{\sim}{x}) = F(x_1, \ldots, x_k) = \int_{\underset{\sim}{y} \leq \underset{\sim}{x}} f(\underset{\sim}{y})\, d\underset{\sim}{y}$$

and for each $r = 0, 1, 2, \ldots,$ let $F_r(\underset{\sim}{x})$ be defined as follows:

$$(B.3.3)\quad F_r(\underset{\sim}{x}) = \prod_{j=1}^{k} \Big[1 - S_1(x_j)\frac{\delta}{\delta x_j} + \ldots + (-1)^r S_r(x_j)\frac{\delta^r}{\delta x_j^r} + (-1)^{r+1}\int_{-\infty}^{x_j} S_r(x_j)\frac{\delta^{r+1}}{\delta x_j^{r+1}}\Big] F(\underset{\sim}{x})$$

where the functions $S_r(x_j)$ of the real variable x_j are defined by (B.1.1). Thus

$$(B.3.4)\quad F_1(\underset{\sim}{x}) = \prod_{j=1}^{k} \Big[1 - S_1(x_j)\frac{\delta}{\delta x_j} + \int_{-\infty}^{x_j} S_1(x_j)\frac{\delta^2}{\delta x_j^2}\Big] F(\underset{\sim}{x})$$

where $F(\underset{\sim}{x})$ is given by (B.3.2).

Theorem B.3.2. Let $f(\underset{\sim}{x})$ be a function on R_k satisfying the conditions (B.3.1). Then

$$(B.3.5)\quad \sum_{\underset{\sim}{n} \le \underset{\sim}{x}} f(\underset{\sim}{n}) = F_r(\underset{\sim}{x})$$

for $r = 1, 2, \ldots$, where $F_r(\underset{\sim}{x})$ is defined by (B.3.3). The summation in (B.3.5) extends over all integer vectors $\underset{\sim}{n}$ such that $\underset{\sim}{n} \le \underset{\sim}{x}$. More generally if A is any measurable set then

$$(B.3.6)\quad \sum_{\underset{\sim}{n} \in A} f(\underset{\sim}{n}) = \int_A dF_r(\underset{\sim}{x})$$

Proof. We shall give the proof for $r = 1$. The proof for any general r is exactly similar. In the first place it follows from (B.2.5) that the relation (B.3.5) is valid for $k = 1$. The proof is then by induction on the dimension k.

Suppose that (B.3.5) is true in R_k. Let $f(x_1, \ldots x_k, x_{k+1}) = f(\underset{\sim}{x}, x_{k+1})$ be a function on R_{k+1} satisfying the conditions (B.3.1). Then since (B.3.5) is assumed to be true in R_k we have

$$(B.3.7) \qquad \sum_{\underset{\sim}{n} \leq \underset{\sim}{x}} f(\underset{\sim}{n}, y) = \prod_{j=1}^{k} \left[1 - S_1(x_j) \frac{\delta}{\delta x_j} + \int_{-\infty}^{x_j} S_1(x_j) \frac{\delta^2}{\delta x_j^2}\right] G(\underset{\sim}{x}, y)$$

where $\underset{\sim}{x}$ denotes the vector $(x_1, \ldots, x_k)$ in R_k and $G(\underset{\sim}{x}, y)$ is defined as follows:

$$(B.3.8) \qquad G(\underset{\sim}{x}, y) = G(x_1, \ldots, x_k, y)$$

$$= \int_{-\infty < \underset{\sim}{z} \leq \underset{\sim}{x}} f(\underset{\sim}{z}, y)\, d\underset{\sim}{z}$$

Then clearly

$$(B.3.9) \qquad \sum_{n_j \leq x_j} f(n_1, \ldots, n_{k+1}) = \sum_{n_{k+1} < x_{k+1}} \sum_{\underset{\sim}{n} \leq \underset{\sim}{x}} f(\underset{\sim}{n}, n_{k+1})$$

$$= \sum_{n_{k+1} \leq x_{k+1}} h(\underset{\sim}{x}, n_{k+1})$$

where

$$(B.3.10) \qquad h(\underset{\sim}{x}, y) = h(x_1, \ldots, x_k, y) = \sum_{\underset{\sim}{n} \leq \underset{\sim}{x}} f(\underset{\sim}{n}, y) .$$

From (B.3.10), (B.3.7) and (B.3.1) it follows that $h(x, y)$ as a function of the real variable y satisfies the conditions (B.2.1). Consequently from (B.2.5) we have

$$(B.3.11) \qquad \sum_{n_{k+1} \leq x_{k+1}} h(\underset{\sim}{x}, n_{k+1}) = \left[1 - S_1(x_{k+1}) \frac{\delta}{\delta x_{k+1}} + \int_{-\infty}^{x_{k+1}} S_1(x_{k+1}) \frac{\delta^2}{\delta x_{k+1}} \right] H(\underset{\sim}{x}, x_{k+1})$$

where

$$(B.3.12) \qquad H(\underset{\sim}{x}, x_{k+1}) = \int_{-\infty}^{x_{k+1}} h(\underset{\sim}{x}, y)\, dy.$$

From (B.3.7), (B.3.10) and (B.3.12) it follows that

$$(B.3.13) \qquad H(\underset{\sim}{x}, x_{k+1}) = \int_{-\infty}^{x_{k+1}} h(\underset{\sim}{x}, y)\, dy$$

$$= \int_{-\infty}^{x_{k+1}} dy \prod_{j=1}^{k} \left[1 - S_1(x_j) \frac{\delta}{\delta x_j} + \int_{-\infty}^{x_j} S_1(x_j) \frac{\delta^2}{\delta x_j^2}\right] G(\underset{\sim}{x}, y)$$

$$= \prod_{j=1}^{k} \left[1 - S_1(x_j) \frac{\delta}{\delta x_j} + \int_{-\infty}^{x_j} S_1(x_j) \frac{\delta^2}{\delta x_j^2}\right] \int_{-\infty}^{x_{k+1}} G(\underset{\sim}{x}, y) dy$$

where $G(x, y)$ is given by (B.3.8). Since

$$\int_{-\infty}^{x_{k+1}} G(\underset{\sim}{x}, y)dy = \int_{-\infty}^{x_{k+1}} dy \int_{-\infty < \underset{\sim}{u} \leq \underset{\sim}{x}} f(\underset{\sim}{u}, y)\, d\underset{\sim}{u}$$

$$= F(x_1, \ldots, x_{k+1})$$

it follows from (B.3.9), (B.3.11) and (B.3.13) that

$$\sum_{n_j \leq x_j} f(n_1, \ldots, n_{k+1}) = \prod_{j=1}^{k+1} \left[1 - S_1(x_j)\frac{\partial}{\partial x_j} + \int_{-\infty}^{x_j} S_1(x_j)\, \frac{\partial^2}{\partial x_j^2}\right] F(\underset{\sim}{x})$$

where

$$F(x_1, \ldots x_{k+1}) = \int_{-\infty < \underset{\sim}{u} \leq \underset{\sim}{x}} f(u_1, \ldots u_{k+1}) du_1 \ldots du_{k+1}.$$

This shows that the relation (B.3.5) if true for k is also true for $k+1$. This proves the assertion (B.3.5) for all k. The relation (B.3.6) is an immediate consequence of (B.3.5). This completes the proof.

The above formula sums the function over the lattice of integer vectors. It is clear, however, that a similar formula can be given (for instance B.2.6 for the case $k = 1$) when the function is summed over a general lattice in R_k. Consider the lattice L of vectors in R_k given by

$$L = \left\{ \underset{\sim}{a} + H\,\underset{\sim}{n},\ \underset{\sim}{n} \text{ - an integer vector} \right\}$$

where H is a nonsingular matrix and $\underset{\sim}{a}$ a fixed vector in R_k. Thus

$$(B.3.14) \qquad \sum_{\underset{\sim}{a}+H\underset{\sim}{n}\ \varepsilon\ A} f(\underset{\sim}{a} + H\underset{\sim}{n}) = \sum_{\underset{\sim}{n}\ \varepsilon\ A} g(n)$$

where $g(\underset{\sim}{x}) = f(\underset{\sim}{a} + H\underset{\sim}{x})$. It is then not difficult to write down for the sum (B.3.14) a closed expression similar to (B.3.5). We confine ourselves to stating the result for one of simplest lattices:

$$L_h = [\underset{\sim}{a} + h\underset{\sim}{n}\ ;\ \underset{\sim}{n} - \text{an integer vector}\ ]$$

where h is a real number.

Theorem B.3.2. Let the function $f(\underset{\sim}{x})$ satisfy the conditions (B.3.1). Then for every integer $r \geq 1$,

$$h^k \sum_{\underset{\sim}{a} + h\underset{\sim}{n} \leq \underset{\sim}{x}} f(\underset{\sim}{a} + h\underset{\sim}{n}) = F_r(\underset{\sim}{x})$$

where

$$F_r(\underset{\sim}{x}) = \prod_{j=1}^{k} \Big\{ 1 - hS_1[(x_j - a_j)/h] \frac{\partial}{\partial x_j} + \ldots + (-1)^r h^r S_r[(x_j - a_j)/h] \frac{\partial^r}{\partial x_j^r}$$
$$+ (-1)^{r+1} h^{r+1} \int_{-\infty}^{x_j} S_r[(x_j - a_j)/h] \frac{\partial^{r+1}}{\partial x_j^{r+1}} \Big\} F(\underset{\sim}{x})$$

where $F(x)$ is defined by (B.3.2).

BIBLIOGRAPHY

[1] Alexandroff, A. D.: Additive set-functions in abstract spaces. Mat. Sbornik, New Series 8, (1940), pp. 307-342; 9, 1941, pp. 563-621; 13, 1943, pp. 169-234.

[2] Bateman, P. T. and others: Seminar on convex sets, conducted at the Institute for advanced study (mimeographed) 1955.

[3] Bergström, H.: On the central limit theorem in R_k, $k > 1$. Skand. Aktuarietidskr, 28, (1945), pp.106-127.

[4] ------ On the central limit theorem in the case of not equally distributed random variables. Skand Aktuarietidskr, 33 (1949), p. 37-62.

[5] Berry, A. C.: The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. 49, (1941), pp. 122-136.

[6] Bochner, S. : Lectures on Fourier Integrals. Translated by Morris Tenenbaum and Harry Pollard. Princeton Univ. Press 1959.

[7] Bonnessen, T. and Fenchel, W.: Theorie der Konvexen Körper. Chelsea Publications 1948.

[8] Blum,J.R. : On the convergence of empirical distribution functions. Ann. Math. Stat., 26 (1955) pp. 527-529.

[9] Cameron, R. H. and Martin, W. T. : Transformations of Wiener integrals under translation. Ann. Math., Series 2, 45 (1944), pp. 386-396.

[10] Cramér, H. : Random variables and probability distributions. Cambridge Tracts No.36, (1937).

[11] Donsker, M. D.: Justification and Extension of Doob's heuristic approach to the Kolmogorov-Theorems., Ann. Math. Stat., 23 (1952) pp. 277-281.

[12] Doob, J. L. : Heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Stat., 20 (1949) pp.393-403.

[13] --------- : Stochastic Processess. John Wiley and sons, New York 1953.

[14] Eggleston, H. G.: Convexity. Cambridge tract No.47, 1958.

[15] Esseen, C. G. : Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta. Math., 77 (1945), pp. 1-125.

[16] Fortet, R. and Mourier, E. : Convergence de la repartition empirique vers la repartition theorique. Annales Scientifiques de l'ecole normale superieure, 60 (1953), pp. 267-285.

[17] Gnedenko, B. V.: On limit theorems of the theory of probability. Address delivered at International Math. Conference Edinghburgh, 1958.

[18] Gnedenko, B. V. and Kolmogorov, A. N. : Limit distributions for sums of independent random variables, Translated by K. L. Chung. Addison Wesley Publishing Company, 1954.

[19] Halmos, P. R. : Lectures on ergodic theory. Published by the mathematical society of Japan 1956.

[20] Hille, E. and Phillips, R. S.: Functional analysis and semi-groups. Volume XXXI of the Amer. Math. Soc. Colloquim publications, 1957.

[21] Kelley, J. L. : General Topology, Van Nostrand and Company, New York, 1955.

[22] Kolmogorov, A. N.: On the Skorohod convergence. Teorya Veryotnostei i yeyaw primanenya 1, (1956), pp. 239-247.

[23] Knopp, K. : Theory and application of infinite series. Translated from the second German Ed. by Miss R. C. Young. Blackie and sons, London, 1928.

[24] Kuratowski, C. : Topologie I. Espaces metrizables; Espaces complets. 2nd Ed. Warszawa - Wroclaw, 1948.

[25] LeCam, L. : Convergence in distribution of stochastic processes. Univ. of California publications in statistics, 2, no. 11, (1957), pp. 207-236.

[26] Lévy, Paul : Processus stochastiques et movement Brownien. Ganthiers-Villars, Paris, 1948.

[27] Loéve, M : Probability Theory. Van Nostrand and Co. New York 1955.

[28] Mourier, E. : Elements aleatoires dans un espace de Banach. Annales de l'institut Henri Poincare, 13 (1953) pp. 161 - 244.

[29] Prohorov, Yu. V. : Convergence of random processes and limit theorems of the theory of probability. Teorya Veryotnostei i yeyau primanenya 1 (1956) pp.177-238.

[30] Prohorov, Yu. V. : Asymptotic behaviour of the binomial distribution. Uspechi. Matem. Nauk, 8 (1953), p.135.

[31] Scheffé, H : A useful convergence theorem for probability distributions. Ann. Math. Stat. 18 (1947), pp. 434-438.

[32] Skorohod, A. V.: Limit theorems for stochastic processes. Teorya Veryotnostei i yeyau primanenya 2 (1957), pp. 289-319.

[33] Skorohod, A. V. : On the differentiability of measures which correspond to stochastic processes. Teorya Veryotnostei i yeyau primanenya 2 (1957), pp. 417-443.

[34] Tucker, H. G.: A generalization of the Glivenko-Cantelli theorem. Ann. Math. Stat., 30 (1959) pp. 828-830.

[35] Varadarajan, V. S. : Convergence in distribution of stochastic processes. Thesis submitted to the University of Calcutta (unpublished) 1958.

[36] Varadarajan, V. S. : Weak convergence of measures on separable metric spaces. Sankhyā 19 (1958), pp. 15-22.

[37] ---------- : On the convergence of sample probability distributions. Sankhyā 19 (1958), pp. 23-26.

[38] Wald, A. : Statistical Decision functions. John Wiley and Sons, New York 1950.

[39] Wallace, D. L.: Asymptotic approximations to distributions. Ann. Math. Stat., 29 (1958), pp. 635-654.

[40] Watson, G. N. : Treatise on Bessel functions, 2nd Edition. Cambridge Univ. Press 1944.

[41] Wolfowitz, J.: Generalization of the theorem of Glivenko-Cantelli. Ann. Math. Stat., 25 (1954), pp. 131-138.